



Universidad Autónoma de Querétaro
Facultad de Ingeniería

Reconstrucción de objetos 3D usando cámaras RGB-D y
Redes Neuronales

Tesis

Que como parte de los requisitos para
obtener el Grado de

Maestro en Ciencias en Inteligencia Artificial

Presenta:

Juan Manuel Aviña Muñoz

Dirigido por:

Dr. Jesús Carlos Pedraza Ortega

Co-Director:

Dr. Juan Manuel Ramos Arreguín

Querétaro, Qro. a Octubre 2024

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



SinDerivadas — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría en Ciencias en Inteligencia Artificial

Reconstrucción de objetos 3D usando cámaras RGB-D y Redes
Neuronales

Tesis

Que como parte de los requisitos para obtener el Grado de
Maestro en Ciencias en Inteligencia Artificial

Presenta
Juan Manuel Aviña Muñoz

Dirigido por:
Dr. Jesús Carlos Pedraza Ortega

Co-dirigido por:
Dr. Juan Manuel Ramos Arreguín

Dr. Jesús Carlos Pedraza Ortega
Presidente

Dr. Juan Manuel Ramos Arreguín
Secretario

Dr. Luis Rogelio Roman Rivera
Vocal

Dr. Efrén Gorrostieta Hurtado
Suplente

MCIA Osmar Antonio Espinosa Bernal
Suplente

Centro Universitario, Querétaro, Qro.
Octubre 2024
México



Dedicatorias

Dedico esta tesis a mi querida familia, cuya constante fé en mí ha sido una fuente inagotable de motivación. A mis padres, por su amor incondicional, apoyo constante y por enseñarme el valor del esfuerzo y la perseverancia.

A Jesús Carlos Pedraza Ortega y a todos los profesores y mentores que han guiado mi formación académica. Su conocimiento, dedicación y pasión por la enseñanza han sido una fuente de inspiración constante. Su orientación y apoyo han sido invaluable en la culminación de esta tesis.

A mis amigos, por su incansable apoyo, por los momentos de distracción que tanto necesitaba y por creer en mí incluso cuando yo dudaba. Ustedes han sido un pilar esencial en mi vida.

A mis compañeros de estudio, por las largas horas de colaboración, discusión y apoyo mutuo. Juntos hemos superado muchos desafíos y celebrado numerosos éxitos. Este logro es tanto suyo como mío.

Finalmente, a todos aquellos que han creído en mí y me han apoyado de diversas maneras, su fe en mis capacidades y su apoyo moral han sido fundamentales para la realización de este trabajo. Este logro es un reflejo de su confianza y aliento.

Juan Manuel Aviña Muñoz

Octubre 2024

Agradecimientos

Quisiera expresar mi sincero agradecimiento a las siguientes personas e instituciones que han contribuido de manera significativa a la realización de este trabajo de investigación.

A mi director de tesis el Dr. Jesús Carlos Pedraza Ortega y a los miembros del sínodo, el Dr. Juan Manuel Ramos Arreguín y el Dr. Luis Rogelio Roman Rivera por su guía experta, paciencia infinita y constante apoyo a lo largo de este proceso, sus valiosos consejos y retroalimentación han sido fundamentales para el éxito de este proyecto.

Agradezco profundamente a todos los profesores y académicos que compartieron su conocimiento, brindaron orientación y ofrecieron comentarios constructivos que enriquecieron este estudio.

Mi más sincero agradecimiento a mi familia por su amor incondicional, apoyo constante y comprensión durante todas las etapas de este viaje académico ya que sin su aliento y sacrificio, este logro no habría sido posible.

A mis amigos y compañeros, quienes estuvieron a mi lado con palabras de aliento, distracciones bienvenidas y momentos de alegría que aliviaron los desafíos del camino hacia la culminación de este proyecto.

A aquellos cuyo trabajo o palabras me han inspirado y motivado a perseguir la excelencia académica, les estoy profundamente agradecido por su influencia positiva en mi desarrollo profesional.

A CONAHCyT y a la Universidad Autónoma de Querétaro por el apoyo brindado para realizar mis estudios de posgrado.

Su apoyo inquebrantable ha sido fundamental para alcanzar este hito académico, y por ello, les estoy profundamente agradecido.

Índice

Índice general

Índice de figuras	VIII
Índice de tablas	X
1 Introducción	1
2 Antecedentes	5
2.1 Sensores RGB-D	6
2.1.1 Comparativa entre sensores RGB-D	9
2.2 Algoritmo ICP	12
2.3 Inteligencia Artificial	13
2.3.1 Transformers	15
2.4 Estado del Arte	18
3 Hipótesis	23
4 Objetivos	25
4.1 Objetivo general	25
4.2 Objetivos Específicos	25

4.2.1	Cronograma	26
5	Metodología	29
5.1	Métodos	30
5.2	Materiales	38
5.2.1	Equipo de cómputo	38
5.2.2	Software	38
5.2.3	Base de datos	39
5.2.4	Algoritmo.	41
5.3	Consideraciones Éticas	50
6	Resultados y discusión	51
6.1	Proyección futura	63
7	Conclusiones	65
	Bibliografía	67
8	Anexos	75

Índice de figuras

1.1	Ejemplo de configuración de cámara. (ResearchGate, 2018).	2
1.2	Flujo de trabajo de <i>SentiSculpt</i> ® (Security Systems S.A.S, 2022).	4
2.1	Sensor RGB Kinect® (Aravind, Ajay, Priyadarshini, 2015).	7
2.2	Sensor RealSense D415 RGB-D (Intel, 2023).	10
2.3	RealSense D415 RGB-D junto con dos objetos de referencia. (Propia, 2023).	11
2.4	Una cara de la nube de puntos de la figura utilizada.	12
2.5	Aplicaciones de la Inteligencia Artificial ® (Sanket, 2018).	15

5.1	Diagrama de bloques del proyecto.	30
5.2	Nube de puntos capturada, sin preprocesamiento.	31
5.3	Muestra de una cara de la captura del objeto preprocesada.	32
5.4	Unión de dos nubes de puntos del objeto.	33
5.5	Reconstrucción mediante ICP.	35
5.6	Objeto reconstruido con ICP.	36
5.7	Objeto reconstruido con ICP.	37
5.8	Proceso de reconstrucción 3D.	38
5.9	3D shapenets: A deep representation for volumetric shapes. (Z. Wu et al., 2015).	40
5.10	Estrategia de agrupación punto-a-nodo.	43
5.11	Transformer geométrico para registro de nubes de puntos.	44
5.12	Modulo de auto-atención geométrica.	45
5.13	Integración de estructuras geométricas.	47
6.1	Captura de profundidad y RBG de izquierda a derecha respectivamente.	53
6.2	Captura RGB-D del sensor RealSense de Intel.	53
6.3	Objeto reconstruido mediante ICP sin submuestreo, 18 capturas.	54
6.4	Nubes de puntos del objeto escaneado previo a la implementación de los métodos.	55
6.5	(a) Método ICP con diferentes niveles de submuestreo y 10 muestras.	56
6.6	Captura de profundidad y RBG de izquierda a derecha respectivamente.	57
6.7	Error en ICP.	59
8.1	Registro de protocolo.	76
8.2	Comité de Ética.	77
8.3	Constancia Nivel de Inglés.	78
8.4	Constancia no adeudo Laboratorios.	79
8.5	Carta de Requisitos.	80
8.6	Portada del Artículo en LMeM.	81
8.7	Dictamen del Artículo en LMeM.	82
8.8	Derechos de Publicación del Artículo en LMeM.	83

Índice de tablas

2.1	Comparativa entre sensores RGB-D comerciales.	9
2.2	Estado del Arte	21
4.1	Cronograma Febrero 2023 - Octubre 2023	26
4.2	Cronograma Noviembre 2023 - Julio 2024	27
5.1	Comparativa de ICP con diferentes porcentajes de submuestreo.	34
5.2	Hardware empleado y breve descripción.	39
5.3	Software empleado y breve descripción.	39
5.4	Resultados RMSE en función de las iteraciones de entrenamiento.	49
6.1	Comparativa entre métodos de reconstrucción de nubes de puntos.	51
6.2	Parámetros para registro de nubes de puntos.	52
6.3	Comparativa de ICP con diferentes porcentajes de submuestreo.	58

Abreviaturas y Siglas

- BERT — Bidirectional Encoder Representations from Transformers
- CAD — Diseño Asistido por Computadora
- CMOS — Complementary Metal Oxide Semiconductor
- CNN — Red Neuronal Convolutacional
- CV — Computer Vision
- DL — Deep Learning
- GPT — Generative Pre-trained Transformers
- IA — Inteligencia Artificial
- ICP — Iterative Closest Point
- LLM — Large Language Models
- ML — Machine Learning
- NLP — Procesamiento de Lenguaje Natural
- OFF — Object File Format
- OPT — Open Pre-trained Transformers
- PLY — Polygon File Format
- RMSE — Error Cuadrático Medio

- RNN — Redes Neuronales Recurrentes
- RN — Redes Neuronales
- Sensor RGB-D — Sensor Red Green Blue Depth
- SDK — Kit de Desarrollo de Software

Resumen

El objetivo de este proyecto es desarrollar un modelo 3D preciso mediante la integración de datos captados por cámaras RGB-D y su procesamiento mediante Redes Neuronales. El modelo 3D resultante, representado como una nube de puntos que combina información de profundidad y color, servirá como representación visual y geométrica de objetos escaneados dentro de escenas estáticas; el proceso de generación de modelos implica la captura de datos tridimensionales mediante cámaras RGB-D, que proporcionan información de profundidad y color; posteriormente, estos datos se procesan mediante algoritmos de Redes Neuronales, aprendiendo y reconociendo patrones en las imágenes capturadas para inferir la geometría subyacente. Dicho proceso permite crear un modelo 3D que retrata con precisión la forma y estructura del objeto. El resultado previsto de este proyecto podría tener un gran impacto en campos como la ingeniería, el diseño y la medicina.

Palabras clave: Reconstrucción 3D, Cámaras RGB-D, Redes Neuronales, Aprendizaje Profundo, Nubes de Puntos, Modelos 3D.

Abstract

The objective of this project is to develop an accurate 3D model by integrating data captured by RGB-D cameras and processing it using Neural Networks. The resulting 3D model, represented as a point cloud combining depth and color information, will serve as a visual and geometric representation of scanned objects within static scenes; the model generation process involves the capture of three-dimensional data by RGB-D cameras, which provide depth and color information; subsequently, this data is processed by Neural Network algorithms, learning and recognizing patterns in the captured images to infer the underlying geometry. Such a process allows the creation of a 3D model that accurately portrays the shape and structure of the object. The expected outcome of this project could have a major impact in fields such as engineering, design and medicine.

Key Words: 3D Reconstruction, RGB-D Cameras, Neural Networks, Deep Learning, Point Clouds, 3D Models.

Capítulo 1

Introducción

La reconstrucción de objetos 3D, esa increíble habilidad de dar vida a objetos del pasado o de plasmar ideas en modelos digitales, no es un invento reciente. Su historia se remonta a tiempos remotos, donde la creatividad y el ingenio humano se unían a las herramientas disponibles para dar forma a la visión.

Las primeras técnicas de reconstrucción 3D se basaban en métodos manuales y artesanales. En la antigua Grecia, por ejemplo, los escultores tallaban a mano representaciones tridimensionales de dioses, héroes y figuras mitológicas a partir de bloques de mármol o bronce. La precisión y el detalle de estas obras, como las esculturas del Partenón o las estatuas romanas, son un claro ejemplo de la maestría alcanzada en la era preindustrial [1].

Con la llegada de la fotografía en el siglo *XIX*, la reconstrucción de objetos 3D dio un salto importante. Las fotografías proporcionaban una base precisa para crear modelos tridimensionales, permitiendo a arquitectos, ingenieros y artistas obtener representaciones más fieles de la realidad. Técnicas como la fotogrametría, que utiliza fotografías tomadas desde diferentes ángulos para calcular las dimensiones de un objeto, se convirtieron en herramientas esenciales [2].

La historia de la correspondencia en geometría 3D se remonta a los primeros días de los gráficos y la visión por computadora. Con el avance de las tecnologías de detección a distancia, esta investigación se ha vuelto esencial para diversas aplicaciones, como la estimación de la pose de objetos, la recuperación y reconstrucción 3D de objetos, así como la localización de cámara [3] mostrado en la Figura 1.1.

En el ámbito de la visión por computadora y la reconstrucción 3D, el registro preciso de nubes de puntos es un reto crucial que impacta directamente en la calidad y utilidad de los modelos tridimensionales generados. Este proyecto se centra en la evaluación de métodos iterativos, específicamente el empleo de CNN's, para el registro efectivo de nubes de puntos. Las CNN han demostrado ser herramientas potentes para el aprendizaje de características y la resolución de tareas complejas en visión artificial, ofreciendo un enfoque alternativo al ICP en el registro de nubes de puntos [4], [5], [6].

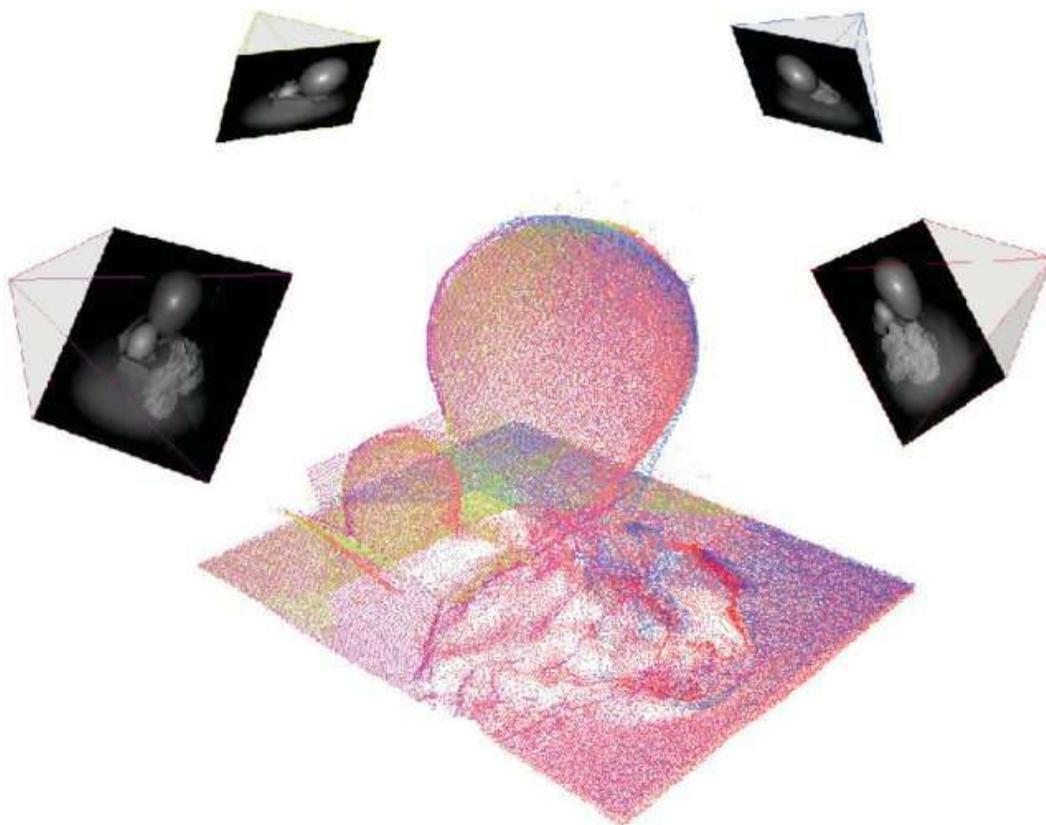


Figura 1.1: Ejemplo de configuración de cámara. (ResearchGate, 2018).

El estudio se centró en la captura de un objeto utilizando un sensor RGB-D, ampliamente reconocido por su capacidad para generar datos detallados tanto en color como en profundidad. La elección de este sensor es crucial, ya que proporciona una rica fuente de datos que es esencial para el proceso de unión de nubes de puntos. La metodología empleada implicó la captura de múltiples vistas del objeto, seguida de la aplicación del método de CNN seleccionado para alinear y fusionar estas vistas en una representación 3D cohesiva y detallada

[7], [8], [5]. Con el desarrollo de este proyecto, se busca generar conocimiento y avances significativos en la reconstrucción 3D de escenas estáticas, es decir, aquellas donde el objeto escaneado no tiene movimiento durante el proceso de captura de datos.

El enfoque en escenas estáticas permite abordar de manera específica y detallada la tarea de reconstruir objetos y entornos inmóviles. Al utilizar cámaras RGB-D y Métodos Iterativos, se espera obtener resultados altamente precisos y realistas en la representación tridimensional de los objetos.

Al centrarse en la reconstrucción 3D de escenas estáticas, este proyecto tiene el potencial de enriquecer el conocimiento en esta área y abrir nuevas posibilidades para aplicaciones prácticas en diversos campos que se beneficien de modelos 3D precisos y realistas.

También, se ha visto una nueva rama en la utilización de estas tecnologías, la optimización del CAD, permitiendo a los diseñadores en el área de medicina regenerativa a predecir las propiedades mecánicas de innovadores andamiajes, así como la creación de meta-materiales [9], [10], [11].

A través de este análisis, buscamos no solo evaluar la efectividad del método CNN en el contexto del sensor RGB-D, sino también proporcionar información sobre las limitaciones y potenciales mejoras para este proceso. Este estudio es particularmente relevante en aplicaciones que requieren reconstrucciones 3D precisas y de alta resolución, como la realidad aumentada, la conservación del patrimonio cultural, y la robótica, entre otras [12].

La problemática a resolver se centra en la reconstrucción 3D de objetos estáticos utilizando cámaras RGB-D y algoritmos de procesamiento de datos para crear una representación lógica del objeto en forma de nube de puntos. El objetivo principal es obtener un modelo tridimensional preciso y realista del objeto sin necesidad de intervenir físicamente o alterar el entorno circundante, esto debido a que la mayoría de los métodos actuales involucran alteraciones en el entorno, lo que puede limitar la precisión y realismo de los modelos obtenidos.

Al utilizar cámaras RGB-D, se capturan datos que incluyen información de color y profundidad de la escena. Estos datos, combinados con algoritmos de procesamiento, permiten generar una nube de puntos que representa la geometría y la posición espacial de los puntos clave del objeto o la escena escaneada, como el mostrado en la Figura 1.2. El desafío radica en desarrollar algoritmos eficientes y precisos que puedan manejar la gran cantidad de da-

tos recopilados por las cámaras y convertirlos en una nube de puntos coherente y detallada. Además, se deben considerar aspectos como la calidad de los datos capturados, la corrección de errores o artefactos, y la precisión en la alineación y reconstrucción de los puntos para obtener una representación tridimensional confiable.

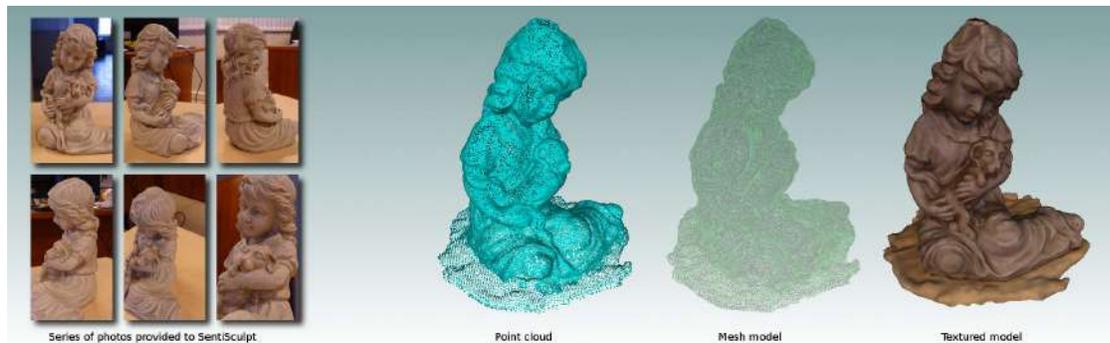


Figura 1.2: Flujo de trabajo de *SentiSculpt*® (Security Systems S.A.S, 2022).

El enfoque de crear el modelo lógico en 3D a partir de datos obtenidos por las cámaras permite realizar la reconstrucción sin poner en riesgo a personas, animales, plantas o el medio ambiente, ya que no requiere interacción física con el objeto ni alteración del entorno. Esto lo convierte en una solución segura y ética para diversas aplicaciones, como el diseño industrial, la arquitectura, la animación y la conservación del patrimonio, entre otros. Sin embargo, la complejidad del problema reside en desarrollar algoritmos avanzados que garanticen la precisión y fidelidad del modelo 3D generado para su utilidad en aplicaciones prácticas y científicas.

Por lo tanto, el algoritmo propuesto tiene como objetivo reducir los tiempos de entrega y minimizar la necesidad de retrabajo en el proceso de diseño y producción de prótesis.

Capítulo 2

Antecedentes

“La inteligencia artificial no tiene que ser buena, solo tiene que ser mejor que los humanos para marcar una diferencia.” – Elon Musk.

Durante décadas, la reconstrucción de objetos en 3D ha representado un desafío constante en el campo de la visión por computadora. Una de las técnicas más efectivas para esta tarea es el uso de Redes Neuronales. En este ensayo, se presentarán algunos antecedentes de la reconstrucción de objetos 3D por medio de redes neuronales.

En la década de 1990, se iniciaron investigaciones para la reconstrucción de objetos tridimensionales con el uso de redes neuronales. Uno de los primeros enfoques, implicó el entrenamiento de una red neuronal para inferir la profundidad a partir de imágenes bidimensionales [13]. Este enfoque fue utilizado en aplicaciones como la conducción autónoma y la navegación de robots.

En la década de 2000, surgieron nuevas técnicas que permitieron la reconstrucción de objetos 3D a partir de imágenes estereoscópicas; estas técnicas implicaban la correspondencia de puntos entre las dos imágenes, lo que permitía la inferencia de la profundidad en cada punto. A partir de estos datos, se podían generar nubes de puntos 3D. Sin embargo, este enfoque tenía limitaciones en cuanto a la precisión y resolución de la reconstrucción [14].

Además, se han desarrollado técnicas para la reconstrucción de objetos 3D a partir de imágenes de escaneo láser y de tomografía computarizada; dichas técnicas permiten la re-

construcción de estructuras complejas, como órganos y tejidos biológicos, y son de gran utilidad en campos como la medicina y la biología [15].

Vivimos en una época donde los avances científicos y tecnológicos avanzan a pasos agigantados, tal que cosas que hace años parecían salidas de la ciencia ficción, hoy en día las tenemos en la palma de nuestras manos. Durante la última década, la Inteligencia Artificial ha sido ampliamente aplicada en diversos campos científicos y en dispositivos cotidianos de nuestros hogares, como teléfonos móviles e incluso aspiradoras. La presencia y el impacto de la Inteligencia Artificial en diferentes áreas de la vida se han vuelto cada vez más notorios y han dado lugar a avances significativos en la tecnología y la forma en que interactuamos con ella [16].

En las últimas décadas, la reconstrucción de objetos 3D mediante redes neuronales ha experimentado una evolución notable. La introducción de redes neuronales convolucionales, técnicas de escaneo láser y tomografía computarizada, y el uso de redes neuronales generativas adversarias han permitido una mejora significativa en la precisión y resolución de la reconstrucción 3D. Estos avances tienen un potencial extraordinario en campos como la medicina, la biología y la ingeniería, abriendo nuevas posibilidades y aplicaciones para obtener representaciones tridimensionales más precisas y detalladas de objetos y estructuras [17], [18], [19], [20].

2.1. Sensores RGB-D

A lo largo de la historia de la humanidad, las cámaras se han vuelto una herramienta esencial. Desde la invención de la fotografía en el siglo XIX hasta la actualidad digital, las cámaras han experimentado una significativa evolución y se han convertido en una parte fundamental de nuestra vida cotidiana. Su presencia y utilidad han trascendido a lo largo del tiempo, siendo un instrumento insustituible en la captura de momentos, la comunicación visual y la documentación de nuestra realidad [21].

El comienzo de los sensores RGB-D fue a finales de la década de 1990, donde su función principal era la del control de velocidad ya que operaban en tiempo real y tenían una alta confiabilidad para detectar vehículos a altas velocidades (Williams, Farmer, & Feldman,

1999).

Para el año 2007, surge la necesidad de emplear estas cámaras para identificar y reconstruir escenarios de alta calidad para la asociación de objetos, es aquí cuando surge una nueva tecnología, llamada “Cámaras de luz estructurada”, sin embargo, hasta el 2010 con la llegada del Kinect (Figura 2.1) es que esta tecnología llega a nuestros hogares de manera relativamente económica; el cual cuenta con una cámara RGB, sensor de profundidad y combinado con un sensor CMOS monocromático, que en conjunto permite la captura de movimiento de cuerpo completo en 3D y reconocimiento facial en cualquier condición lumínica ambiental [22].

Años después, esta tecnología se popularizó en el área científica en proyectos de visión por computadora, modelado 3D debido a su bajo costo y proveyendo nuevas oportunidades en la interacción humano-máquina [23], [24].

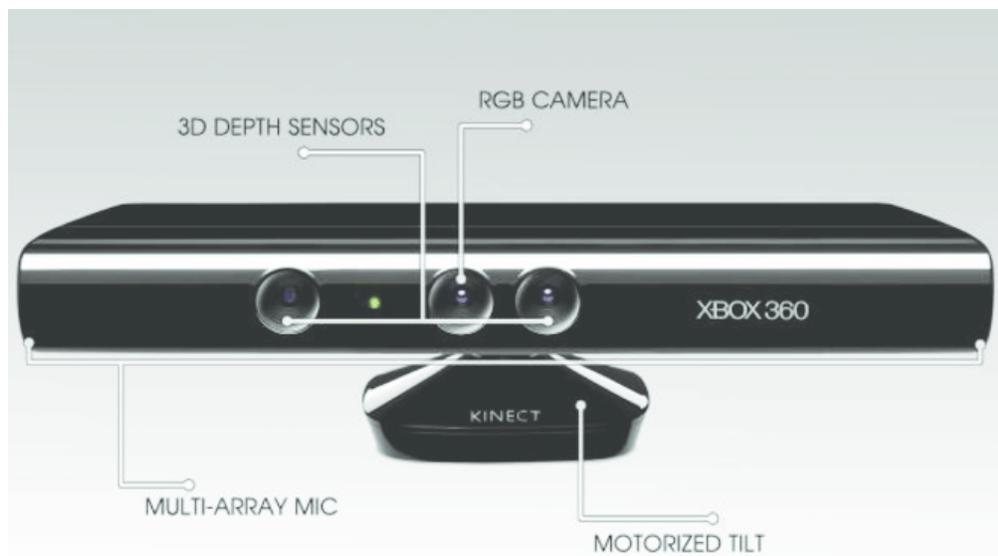


Figura 2.1: Sensor RGB Kinect® (Aravind, Ajay, Priyadarshini, 2015).

Los sensores RGB-D, también conocidos como sensores de profundidad, son dispositivos que combinan una cámara RGB (Red, Green, Blue) con un sensor de profundidad (Depth). Estos sensores son capaces de capturar tanto información de color como información tridimensional del entorno.

La cámara RGB captura imágenes en color, similar a una cámara convencional, y es capaz

de registrar la intensidad de la luz en tres canales: rojo, verde y azul; esta información permite capturar la apariencia visual del entorno, como los colores de los objetos y la iluminación. El sensor de profundidad, por otro lado, utiliza diferentes tecnologías para calcular la distancia entre el sensor y los objetos en la escena [25].

Hay varias tecnologías utilizadas en los sensores RGB-D, incluyendo la luz estructurada, el tiempo de vuelo y el escaneo láser, dichas tecnologías emiten luz infrarroja o láser y miden el tiempo que tarda en rebotar y volver al sensor, lo que proporciona información sobre la distancia [26].

Al combinar la información de la cámara RGB y el sensor de profundidad, se obtiene una representación tridimensional del entorno en tiempo real. Esta representación se conoce como nube de puntos, donde cada punto contiene coordenadas espaciales (x , y , z) y valores de color (rojo, verde, azul) [7].

Los sensores RGB-D ofrecen una amplia gama de aplicaciones beneficiosas en diversos campos, por ejemplo, en la robótica, desempeñan un papel fundamental en la navegación y percepción del entorno, permitiendo a los robots detectar y evitar obstáculos de manera eficiente. Estas capacidades son cruciales para mejorar la autonomía y seguridad de los robots en diferentes entornos, facilitando su interacción y colaboración con el mundo que los rodea [27].

Dentro del ámbito de la realidad virtual y aumentada, los sensores RGB-D desempeñan un papel fundamental al permitir el seguimiento preciso de movimientos y brindar una experiencia inmersiva al interactuar con objetos virtuales. Además, se aplican en la captura de movimiento, facilitando la captura precisa de los movimientos humanos para su posterior reproducción y análisis. Otros usos incluyen la reconstrucción tridimensional de escenas, que permite generar modelos virtuales detallados del entorno real, así como la monitorización de espacios en diversas áreas de aplicación. Los sensores RGB-D ofrecen posibilidades versátiles y prometedoras en estos y muchos otros campos [28], [29].

En el contexto de visión por computadora, las cámaras RGB-D son ampliamente utilizadas en diversas aplicaciones, como la segmentación precisa de objetos, la reconstrucción tridimensional de escenas y la detección de personas y gestos. Las aplicaciones mencionadas aprovechan la valiosa información de profundidad proporcionada por los sensores RGB-D,

permitiendo a los algoritmos de visión por computadora comprender con precisión la posición y geometría de los objetos en escena. La disponibilidad de esta información adicional convierte a las cámaras RGB-D en una herramienta esencial para mejorar la precisión y capacidad analítica en tareas relacionadas con la visión por computadora [30].

Su capacidad para capturar datos de color y profundidad en conjunto potencia significativamente la percepción y toma de decisiones en aplicaciones como la navegación robótica, la realidad virtual y aumentada, así como la detección y seguimiento de objetos en entornos complejos [28].

2.1.1. Comparativa entre sensores RGB-D

Es importante tener en cuenta que existen diferentes marcas y modelos de sensores RGB-D en el mercado, como el Kinect de Microsoft (Figura 2.1), el RealSense de Intel (Figura 2.2) o el ZED de Stereolabs. Cada uno de ellos puede tener características y capacidades ligeramente diferentes, pero en general comparten el objetivo de capturar información visual y de profundidad en un solo dispositivo.

La Tabla 2.1 muestra comparativa entre los sensores RGB-D Kinect de Microsoft, RealSense de Intel y ZED de Stereolabs [31]. Esta tabla proporciona una visión general de algunas de las características clave de cada sensor:

Tabla 2.1: Comparativa entre sensores RGB-D comerciales.

Características	Kinect v2	RealSense D415	ZED mini
Fabricante	Microsoft	Intel	Stereolabs
Tecnología	Tiempo de vuelo	Luz estructurada	Estéreo Visual-inercial
Resolución RGB	1920*1080	1920*1080	2208*1242
Resolución Profundidad	512*424	1280*720	2208x1242
Campo de visión Horizontal	70°	65°	90°
Campo de visión Vertical	60°	40°	60°
Rango de profundidad	0.5 a 4.5 m	0.3 a 4 m	0.10 a 15 m
Precisión de profundidad	1.5 mm @ 50 cm 5 cm @ 4.5 m	<2% @ 2m	<1.5% @ 3m <7% @ 15m
Frecuencia de muestreo	30 fps	90 fps	100 fps
Plataformas soportadas	Windows, Linux, Xbox	Windows, Linux	Windows, Linux
Aplicaciones comunes	Juegos, robótica, realidad virtual.	Robótica, realidad virtual, monitoreo, reconocimiento.	Robótica, realidad virtual, mapeo 3D, robótica.

El sensor RealSense D415 utiliza la tecnología de luz estructurada para calcular la distan-

cia entre el sensor y los objetos en la escena, generando así una nube de puntos que representa la estructura tridimensional del entorno, permitiendo que la cámara sea capaz de comprender la geometría y la posición de los objetos en relación con el sensor, Figura 2.2.

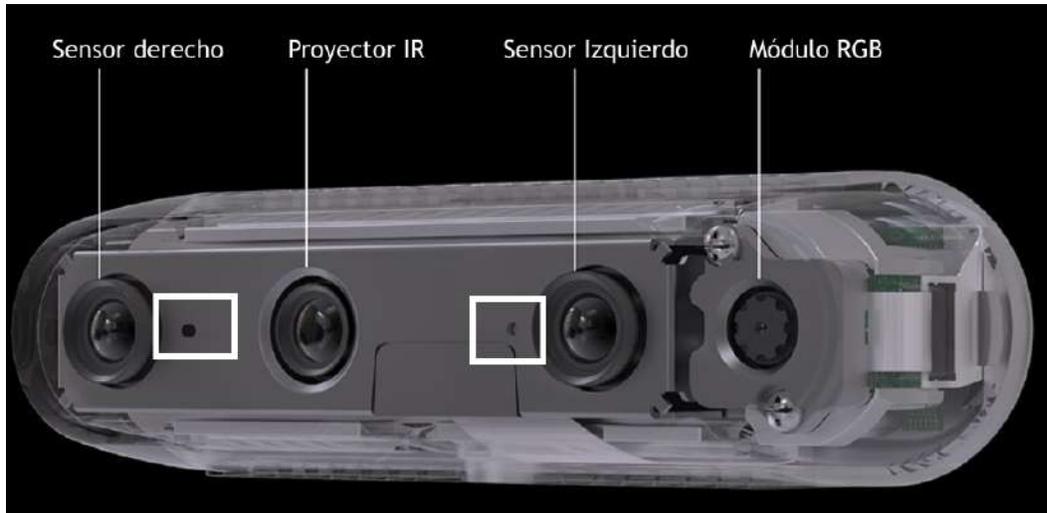


Figura 2.2: Sensor RealSense D415 RGB-D (Intel, 2023).

Una característica distintiva del RealSense D415 es su capacidad de capturar datos de profundidad en tiempo real, lo que la hace ideal para aplicaciones que requieren respuestas en tiempo real, como la navegación de robots o la interacción en entornos de realidad virtual.

Además de su capacidad para capturar imágenes y datos de profundidad, el RealSense D415 es compatible con el SDK de Intel RealSense, que proporciona una serie de herramientas y bibliotecas para desarrolladores. Esto permite a los programadores crear aplicaciones personalizadas que aprovechen las capacidades de la cámara, como la detección de objetos, el seguimiento de movimientos y la reconstrucción 3D [32].

La elección de utilizar la cámara Intel RealSense D415, Figura 2.3 se basa en una evaluación de varios parámetros clave que son cruciales para el éxito del proyecto. A continuación, se analizan estos parámetros y se explica por qué la D415 es una opción adecuada:

- **Portabilidad:** es relativamente compacta y liviana, lo que la hace portátil y adecuada para aplicaciones en una variedad de entornos. Su tamaño reducido facilita la integración en dispositivos móviles, robots y sistemas que requieren movilidad.
- **Tecnología:** la tecnología de tiempo de vuelo utilizada en la D415 es avanzada y con-



Figura 2.3: RealSense D415 RGB-D junto con dos objetos de referencia. (Propia, 2023).

fiable para la captura precisa de datos de profundidad. Esto garantiza una percepción tridimensional confiable y de alta calidad en diferentes escenarios, lo que es esencial para aplicaciones que requieren una representación precisa del entorno.

- **Resolución:** ofrece una resolución de imagen y profundidad adecuada para muchas aplicaciones. Con una combinación de imágenes RGB y datos de profundidad, puede capturar detalles necesarios para el análisis tridimensional y la identificación de objetos en la escena.
- **Costo:** en comparación con algunas soluciones de cámaras de profundidad más avanzadas, el RealSense D415 ofrece un buen equilibrio entre costo y funcionalidad, lo cual lo convierte en una opción atractiva para proyectos con restricciones presupuestarias pero que aún requieren capacidades de percepción 3D.
- **Disponibilidad:** es ampliamente reconocida y respaldada por Intel, asegurando soporte confiable, actualizaciones de firmware y acceso a herramientas de desarrollo, facilitan-

do la creación y mantenimiento de aplicaciones basadas en esta cámara.

2.2. Algoritmo ICP

El algoritmo ICP es un método para alinear modelos 3D o nubes de puntos; la idea básica es minimizar la distancia entre los puntos de dos conjuntos de datos que se corresponden entre sí. Es esencial para la reconstrucción precisa de objetos o escenas en 3D. Este algoritmo itera sobre dos conjuntos de datos (nubes de puntos), buscando minimizar la diferencia entre ellos, comparando puntos de un conjunto con el conjunto más cercano en el otro y calcula la mejor transformación (rotación y traslación) que alinea los dos conjuntos [33], [34], [35].

Este algoritmo es crucial para alinear y fusionar múltiples imágenes o nubes de puntos capturadas desde diferentes ángulos, creando así una representación 3D completa y precisa.

El método implementado incorpora la visualización de una nube de puntos, ilustrada en la Figura 2.4.

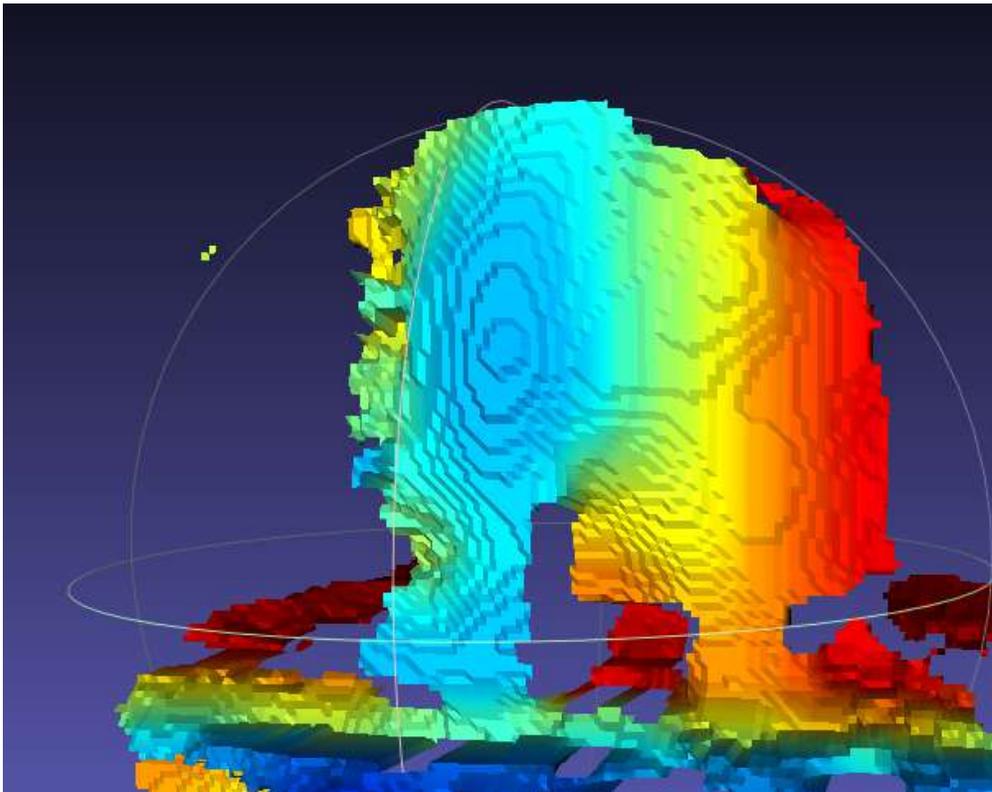


Figura 2.4: Una cara de la nube de puntos de la figura utilizada.

Iteración: El proceso de asignación de puntos más cercanos y minimización del error se repite varias veces; en cada iteración, se recalculan las correspondencias de los puntos más cercanos utilizando la transformación encontrada en la iteración anterior. Esto se hace hasta que el proceso converge, es decir, hasta que el cambio en el error entre iteraciones sucesivas sea suficientemente pequeño.

Error Cuadrático Medio (RMS) : El Error Cuadrático Medio (RMS, por sus siglas en inglés) se utiliza para medir la diferencia entre los valores predichos y los valores reales en un conjunto de datos. La fórmula es la siguiente:

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (2.1)$$

Donde:

- n es el número total de observaciones en el conjunto de datos.
- Y_i representa el valor real de la observación i .
- \hat{Y}_i representa el valor predicho para la observación i .
- La expresión $(Y_i - \hat{Y}_i)^2$ calcula el cuadrado de la diferencia entre el valor real y el valor predicho.

Evaluación de la Convergencia: Se evalúa si el algoritmo ha convergido verificando si el cambio en la métrica de error entre iteraciones consecutivas está por debajo de un umbral específico, en nuestro caso, comparando con el RMS.

2.3. Inteligencia Artificial

La Inteligencia Artificial es una disciplina de ciencias de la computación que busca replicar la inteligencia y sus procesos en las computadoras. Su objetivo central es diseñar sistemas robóticos capaces de tomar decisiones y resolver problemas de manera autónoma. A través de algoritmos avanzados y modelos de aprendizaje automático, la IA permite que las máquinas aprendan y se adapten a nuevas situaciones, mejorando su capacidad para abordar tareas

complejas y tomar decisiones inteligentes en diversos campos de aplicación. Esta área multidisciplinaria fusiona la computación, las matemáticas, la biología y la lógica para diseñar sistemas artificiales que puedan realizar tareas de forma autónoma o con mínima intervención humana, imitando así el comportamiento humano. Esta transformación en la tecnología ha revolucionado la forma en que se abordan los desafíos industriales, permitiendo que la tecnología exhiba un “comportamiento inteligente” y ofrezca soluciones flexibles y confiables para tareas que antes eran exclusivas de los humanos [36].

El uso de la IA ha generado un cambio de paradigmas en la industria, promovido por una innovadora forma de interacción entre humano y máquina; dando forma a lo que hoy conocemos como industria 4.0, caracterizada por fábricas inteligentes en donde humanos y sistemas cibernéticos interactúan en tiempo real, permitiendo un muy alto nivel de automatización en cada uno de los procesos industriales y ayudando a generar un mayor rendimiento en el área donde sean utilizadas; siendo su objetivo principal manipular de manera independiente y coordinada otros subsistemas con múltiples procesos [37], [38] [39].

En la era tecnológica actual, integrar la IA con varias industrias ha sido uno de los factores más importantes para analizar y mejorar los productos fabricados. Simular las máquinas es de vital importancia al momento de diseñar la maquinaria, para de esta manera evitar un gasto innecesario de materia prima, así como de tiempo de trabajo y así disminuir la posibilidad de errores de precisión en el diseño. Debido a la complejidad de las estructuras de la máquina en cuestión, se suelen integrar métodos de IA a los modelos reales, seguidos de un aumento de datos y extracción de características de la máquina real y por último, realizando un análisis exhaustivo de los modelos [40].

Los datos son el componente central de un sistema de IA; esta se divide en dos tipos:

- **Datos estructurados** como archivos de Excel, datos en forma tabular, etc.
- **Datos no estructurados** como emails, PDF's, imagenes, archivos de audio y video, etc.

La identificación, recolección y almacenamiento de los datos juegan un papel importante en el resultado de un sistema de IA [41].

Dentro del amplio campo de la Inteligencia Artificial, actualmente se reconocen cuatro

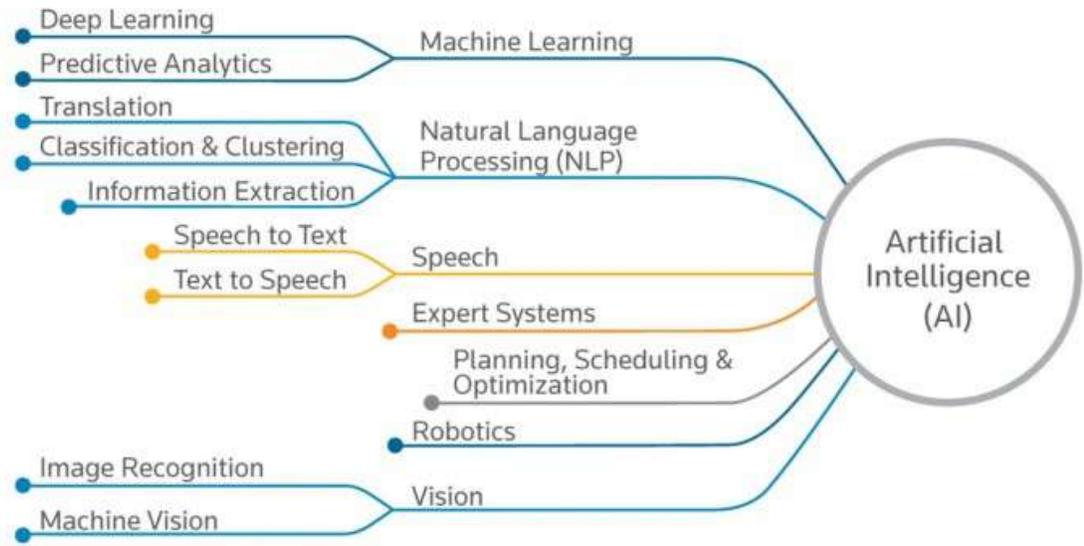


Figura 2.5: Aplicaciones de la Inteligencia Artificial ® (Sanket, 2018).

enfoques principales. Dos de ellos se centran en la imitación de la inteligencia humana, uno busca crear sistemas que piensen como humanos, mientras que el otro busca sistemas que actúen como humanos. Por otro lado, existen enfoques centrados en la racionalidad, con sistemas que piensan de manera racional y sistemas que actúan de manera racional. Estos enfoques reflejan la diversidad de objetivos y métodos que se emplean en la AI, abriendo un abanico de posibilidades para la creación y mejora de sistemas inteligentes que se adaptan a diferentes contextos y tareas [42].

Los modelos de IA aprenden a partir de los datos proporcionados como entrada, y si dichos datos contienen sesgos inherentes, el modelo generará resultados igualmente sesgados [43]. Esto puede derivar en salidas discriminatorias y poco confiables, afectando negativamente tanto la precisión de los resultados como la reputación de la marca asociada.

2.3.1. Transformers

Los Transformers irrumpieron en el mundo del NLP en 2017 con la publicación del artículo “Attention is All You Need” de Vaswani et al [44]. Este trabajo introdujo una arquitectura de red neuronal novedosa que abandonaba las RNN, que eran el estándar en ese momento, y apostaba por un mecanismo de atención basado en capas de autoatención.

Las RNN's, a pesar de su éxito en tareas como la traducción automática y el reconocimiento de voz, tenían algunas limitaciones:

- **Dificultad para manejar dependencias de largo alcance:** Las RNN's procesan la información de forma secuencial, lo que les dificulta capturar relaciones entre palabras que están muy separadas en una oración.
- **Entrenamiento lento:** El procesamiento secuencial de las RNN's las hace más lentas de entrenar que los modelos que pueden procesar la información en paralelo.

Los Transformers abordan estas limitaciones utilizando dos mecanismos clave:

- **Atención:** La atención permite que el modelo se centre en las partes más relevantes de la entrada al procesar cada palabra. Esto le permite capturar dependencias de largo alcance de manera más efectiva que las RNN. El mecanismo de atención se basa en la siguiente fórmula:

$$attention(Q, K, V) = softmax(Q * K^T) * V \quad (2.2)$$

Donde:

- Q, K y V son matrices de dimensiones $(d_{model}, num_{heads}, seq_{len})$
- d_{model} es la dimensión del modelo
- num_{heads} es el número de cabezas de atención
- seq_{len} es la longitud de la secuencia
- T es la transpuesta

La función *attention* toma tres matrices como entrada: Q (*query*), K (*key*) y V (*value*). Las matrices Q y K se calculan a partir de la entrada, mientras que la matriz V contiene las representaciones de las palabras de la entrada. La función *attention* calcula una matriz de pesos de atención, donde cada peso indica la importancia de una palabra en particular para la palabra que se está procesando actualmente. La matriz de pesos de atención se utiliza luego para ponderar los vectores en la matriz V , lo que produce una representación contextual de la palabra actual [45].

- **Procesamiento en paralelo:** Los Transformers pueden procesar la información en paralelo, lo que los hace mucho más rápidos de entrenar que las RNNs. Esto se debe a que el mecanismo de atención no depende del orden de las palabras en la secuencia, lo que permite procesar todas las palabras de forma simultánea.
- **Función de escalado:** La función de escalado se utiliza para ajustar la dimensión de las entradas al mecanismo de atención. Se puede expresar con la siguiente fórmula:

$$scaled_{attention} = \frac{attention}{\sqrt{d_k}} \quad (2.3)$$

Donde:

- $scaled_{attention}$ es la matriz de atención escalada
- $attention$ es la matriz de atención sin escalar
- d_k es la dimensión de la clave (K)

Impacto de los Transformers: Los Transformers han tenido un impacto profundo en el NLP, impulsando avances en una amplia gama de tareas, incluyendo:

- Los modelos basados en la arquitectura de Transformers han demostrado un rendimiento superior a los modelos tradicionales basados en RNNs en casi todas las métricas relacionadas con la traducción automática. Un claro ejemplo es el sistema Google Translate, el cual ha mejorado notablemente la precisión de las traducciones entre diferentes pares de idiomas, gracias a la capacidad de los Transformers para manejar dependencias de largo alcance y procesar datos de manera eficiente en paralelo [46].
- En el ámbito de la generación de resúmenes de texto, los Transformers también han superado a los enfoques anteriores. El modelo de Google AI Text Summarization es un ejemplo representativo de esta evolución, logrando generar resúmenes más precisos y coherentes a partir de artículos extensos, como noticias y documentos largos, de forma completamente automática [47].

- Además, los Transformers han permitido avances significativos en la capacidad de los modelos para responder preguntas complejas de forma más completa y contextualizada. Un ejemplo relevante es LaMDA de Google AI, un modelo conversacional que puede interactuar de manera fluida y responder a preguntas abiertas, desafiantes o incluso inusuales, proporcionando respuestas más informativas que los modelos tradicionales [48].
- Los Transformers también han sido aplicados con éxito a la generación de texto de alta calidad en diferentes contextos, como la creación de poemas, código y guiones. Un ejemplo destacado es GPT-3 de OpenAI, el cual ha demostrado una capacidad sin precedentes para generar texto coherente y creativo en una variedad de géneros y estilos [49].

Más allá del NLP: Los Transformers no solo han revolucionado el campo del Procesamiento del Lenguaje Natural (NLP), sino que también han demostrado su utilidad en otros dominios de la inteligencia artificial, como la visión artificial y el procesamiento del habla. La arquitectura de los Transformers, con su capacidad para procesar información en paralelo y capturar dependencias de largo alcance, se ha convertido en una herramienta versátil y poderosa para una amplia gama de tareas.

En visión artificial, los Transformers se han empleado con éxito para el desarrollo de modelos que superan a las arquitecturas anteriores en la detección de objetos. Un ejemplo sobresaliente es DETR de Google AI, que utiliza la arquitectura de Transformers para detectar objetos en imágenes en tiempo real, logrando una precisión notable y un procesamiento eficiente [50].

2.4. Estado del Arte

La implementación de algoritmos para la generación y optimización de objetos 3D ha permitido a los diseñadores y fabricantes de prótesis crear sistemas con configuraciones adaptables. Esto significa que los diseños de prótesis pueden ser personalizados para adaptarse a la zona específica del cuerpo en la que se aplicará la parte generada, ya sea en piernas, brazos,

dedos, etc. Además, también se pueden tener en cuenta las cargas a las que se someterá la pieza conforme a los requerimientos de cada paciente.

El ML es una disciplina de la IA enfocada en desarrollar modelos y algoritmos capaces de aprender de datos y hacer predicciones o tomar decisiones basadas en ellos. En los últimos años, el ML ha crecido de manera exponencial gracias a los avances en tecnologías de cómputo, la accesibilidad a grandes cantidades de datos y la aparición de nuevas técnicas de aprendizaje profundo [42].

En la actualidad, el ML se aplica de manera extensa en diversos campos, desde la detección de fraudes en transacciones financieras hasta la creación de diagnósticos médicos, pasando por la conducción autónoma de vehículos, la optimización de procesos industriales y la personalización de servicios en línea. Su versatilidad lo ha convertido en una herramienta fundamental para mejorar la eficiencia y eficacia de numerosas empresas y organizaciones, permitiendo tomar decisiones más informadas y precisas, así como ofreciendo soluciones innovadoras y adaptadas a diferentes necesidades y sectores [51].

En cuanto a las técnicas de ML, se pueden clasificar en dos grandes grupos: aprendizaje supervisado y aprendizaje no supervisado. El primero se basa en entrenar el modelo con un conjunto de datos etiquetados, es decir, datos en los que se conoce la respuesta correcta, para que pueda predecir la salida correspondiente a nuevos datos. El segundo, en cambio, se utiliza para encontrar patrones y estructuras en un conjunto de datos no etiquetados, sin necesidad de conocer la respuesta correcta de antemano.

Entre las técnicas más utilizadas en el aprendizaje supervisado se encuentran las redes neuronales, el árbol de decisiones, la regresión lineal y los métodos de clasificación, mientras que en el aprendizaje no supervisado se destacan el clustering, la reducción de dimensionalidad y las redes neuronales auto-encoder.

En cuanto a las herramientas y frameworks utilizados en ML, los más populares son TensorFlow, PyTorch, Scikit-Learn y Keras, entre otros; dichas herramientas proporcionan una gran cantidad de algoritmos y modelos pre-entrenados, así como herramientas que permiten procesamiento de datos y la visualización de resultados.

También, se ha visto una nueva rama en la utilización de estas tecnologías, la optimización del CAD por sus siglas en inglés, permitiendo a los diseñadores en el área de medicina

regenerativa a predecir las propiedades mecánicas de innovadores andamiajes, así como la creación de meta-materiales [9], [10], [11].

Con el desarrollo de este proyecto, se busca generar conocimiento y avances significativos en la reconstrucción 3D de escenas estáticas, es decir, aquellas donde el objeto escaneado no tiene movimiento durante el proceso de captura de datos. El enfoque en escenas estáticas permite abordar de manera específica y detallada la tarea de reconstruir objetos y entornos inmóviles. Al utilizar cámaras RGB-D y Redes Neuronales, se espera obtener resultados altamente precisos y realistas en la representación tridimensional de los objetos.

Los beneficios de enfocarse en escenas estáticas incluyen:

- **Mayor precisión:** al no tener que lidiar con movimientos en los objetos escaneados, se pueden obtener modelos más precisos y detallados, ya que no hay desplazamientos ni deformaciones temporales que afecten la reconstrucción.
- **Eficiencia en la captura de datos:** al trabajar con objetos estáticos, se puede capturar la información necesaria de manera más rápida y sencilla, lo que contribuye a una mayor eficiencia en el proceso de reconstrucción.
- **Reducción de la complejidad:** al evitar la necesidad de rastrear movimientos, se simplifica el proceso de reconstrucción, lo que puede facilitar la implementación de técnicas avanzadas como las Redes Neuronales para obtener mejores resultados.
- **Aplicaciones en diversos campos:** la reconstrucción 3D de escenas estáticas tiene aplicaciones prácticas en campos como el diseño industrial, la arquitectura, la conservación del patrimonio, la ingeniería inversa, entre otros, donde se requiere una representación tridimensional precisa de objetos y entornos estáticos.

Al centrarse en la reconstrucción 3D de escenas estáticas, este proyecto tiene el potencial de enriquecer el conocimiento en esta área y abrir nuevas posibilidades para aplicaciones prácticas en diversos campos que se beneficien de modelos 3D precisos y realistas.

Gracias a los avances en técnicas, herramientas y datos, se espera que continúe creciendo y transformando nuestra forma de vida en el futuro.

Tabla 2.2: Estado del Arte

Autor	Método y Precisión	Resultados Clave
Qi et al. (2017)	PointNet 89.2 %	Se provee un modelo que unifica tareas de reconocimiento 3D obteniendo mejores resultados que el estado del arte.
Maturana et al. (2018)	VoxNet 92 %	Reconocimiento y clasificación de objetos en tiempo real.
Smith et al. (2020)	CNN 86 %	Mayor precisión en la reconstrucción 3D de objetos estáticos y mejora significativa en la calidad visual de los modelos.
Wang et al. (2022)	RSNet++ 99.2 %	Reducción del ruido en los modelos 3D y aumento de la precisión en la representación geométrica.
Azinović et al. (2022)	MLP 92.4 %	Generación de mapas 3D más precisos en comparación con métodos tradicionales.

Capítulo 3

Hipótesis

Considerando aspectos cualitativos y cuantitativos, se evidenciará que el uso de sensores RGB-D y Redes Neuronales para la reconstrucción 3D de objetos estáticos resultando en modelos más precisos y realistas comparado con los métodos tradicionales.

Capítulo 4

Objetivos

4.1. Objetivo general

El objetivo principal de este trabajo es crear un sistema de reconstrucción de objetos en tres dimensiones utilizando técnicas de visión por computadora, sensores RGB-D, y técnicas basadas en redes neuronales convolucionales. La implementación y evaluación del sistema son objetivos clave en el trabajo, y la aplicación potencial del sistema es amplia y variada en diferentes campos de aplicación.

4.2. Objetivos Específicos

Los objetivos específicos de este proyecto son:

- Definir e implementar un algoritmo que permita hacer reconstrucción de objetos en tres dimensiones.
- Proponer un algoritmo basado en redes neuronales para mejorar la reconstrucción 3D.
- Generar un modelo en 3D haciendo uso de las técnicas descritas.
- Definir las métricas adecuadas para evaluar el desempeño del algoritmo.
- Comparar los resultados con algún sistema del estado del arte para validación.

4.2.1. Cronograma

El siguiente cronograma incluye las fechas propuestas para el proyecto.

Tabla 4.1: Cronograma Febrero 2023 - Octubre 2023

Actividades	2023								
	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct
Seleccionar base de datos	■	■	■						
Implementación de algoritmos de referencia			■	■	■				
Pruebas de algoritmos de referencia				■	■				
Selección de Arquitectura					■	■			
Implementación del modelo						■	■	■	■
Entrenamiento de algoritmo								■	■
Selección de algoritmo de compresión									
Implementación de algoritmo de compresión									
Pruebas del sistema									
Evaluación del sistema									
Análisis de resultados									
Redacción del artículo científico									
Documentación			■	■	■	■	■	■	■
Redacción de tesis	■	■	■	■	■	■	■	■	■

Tabla 4.2: Cronograma Noviembre 2023 - Julio 2024

Actividades	2023		2024						
	Nov	Dic	Ene	Feb	Mar	Abr	May	Jun	Jul
Seleccionar base de datos									
Implementación de algoritmos de referencia									
Pruebas de algoritmos de referencia									
Selección de Arquitectura									
Implementación del modelo									
Entrenamiento de algoritmo									
Selección de algoritmo de compresión									
Implementación de algoritmo de compresión									
Pruebas del sistema									
Evaluación del sistema									
Análisis de resultados									
Redacción del artículo científico									
Documentación									
Redacción de tesis									

Capítulo 5

Metodología

El proyecto se realizó en la Facultad de Ingeniería Campus Aeropuerto de la Universidad Autónoma de Querétaro, Querétaro.

1. Seleccionar una base de datos pública, apropiada para representar el conjunto de datos para entrenamiento y validación.
2. Investigar y elegir arquitecturas de redes neuronales y aprendizaje profundo que se utilizará para el proyecto.
3. Definir una estrategia de entrenamiento apropiada para la base de datos seleccionada.
4. Realizar el entrenamiento de la Red Neuronal de acuerdo a la estrategia previamente seleccionada.
5. Definir e implementar el algoritmo de compresión basado en la arquitectura de la Red Neuronal.
6. Estudiar y analizar los algoritmos de poda neuronal de forma que no impacten la métrica de precisión.
7. Realizar la etapa de compresión del modelo previamente entrenado.
8. Definir las pruebas de desempeño que se utilizarán para evaluar la Red Neuronal comparado con modelos del estado del arte.

9. Realizar pruebas de desempeño.
10. Analizar e interpretar resultados obtenidos.
11. Presentar resultados finales, conclusión y comparación para comprobar la hipótesis planteada.

5.1. Métodos

Para asegurar el éxito de este proyecto, se implementaron una serie de métodos rigurosos y bien fundamentados, que se detallan a continuación:

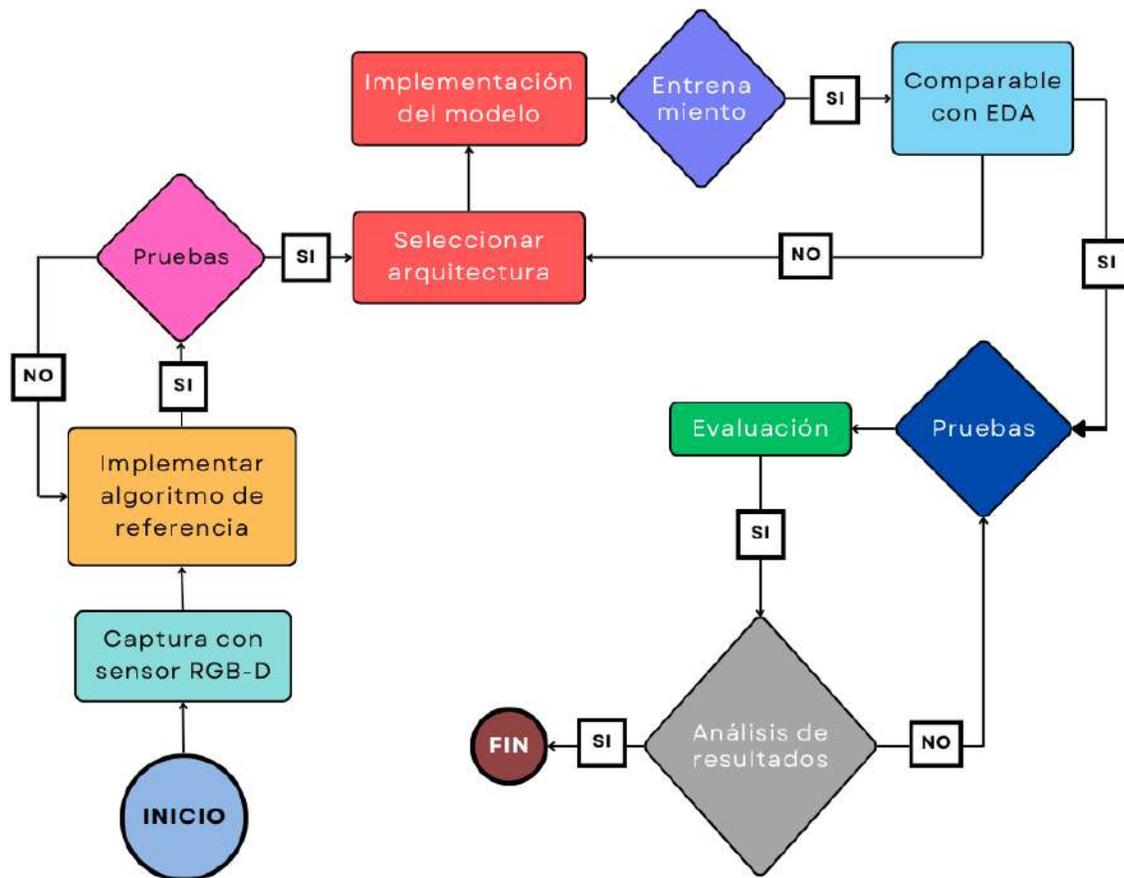


Figura 5.1: Diagrama de bloques del proyecto.

- **Adquisición de datos:** Para llevar a cabo el procedimiento, se empleó el sensor Real-Sense D415. El objeto fue escaneado mediante un barrido completo de 360°, captu-

rando imágenes cada 10° con el uso de una base giratoria y ajustando la profundidad de la captura. En la Figura 5.2 se puede observar que solo la porción en color azul de la escena corresponde a la nube de puntos obtenida. Este enfoque permitió generar un total de 36 capturas, asegurando una cobertura exhaustiva y detallada del objeto desde todos los ángulos posibles. La captura desde múltiples perspectivas tuvo como propósito minimizar la pérdida de información relevante y maximizar la calidad de los datos recopilados para su posterior procesamiento.

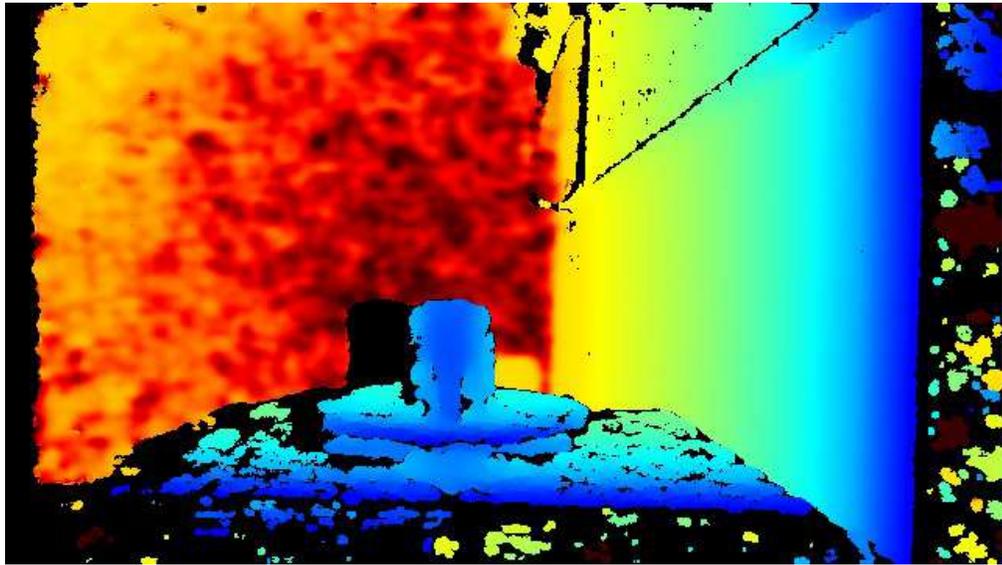


Figura 5.2: Nube de puntos capturada, sin preprocesamiento.

- **Preprocesamiento de datos:** Después de la captura de las nubes de puntos, se llevó a cabo una fase de limpieza con el objetivo de eliminar la mayor cantidad posible de ruido, mostrado en la Figura 5.3, esto fue realizado con ayuda del software MeshLab [52]; este procedimiento de limpieza se realizó de forma manual, es decir, se eliminaron todos los puntos que no fuesen parte del objeto escaneado, como por ejemplo, la base donde se colocó el objeto. Una vez finalizada la limpieza, se realizó un submuestreo con ayuda del software CloudCompare de las tomas capturadas para agilizar el proceso de procesamiento, este submuestreo implicó la reducción del 10 %, 20 %, 30 %, 40 % y 50 % de los puntos originales, esto con la finalidad de hacer una comparativa de precisión y tiempos de ejecución para el método seleccionado.

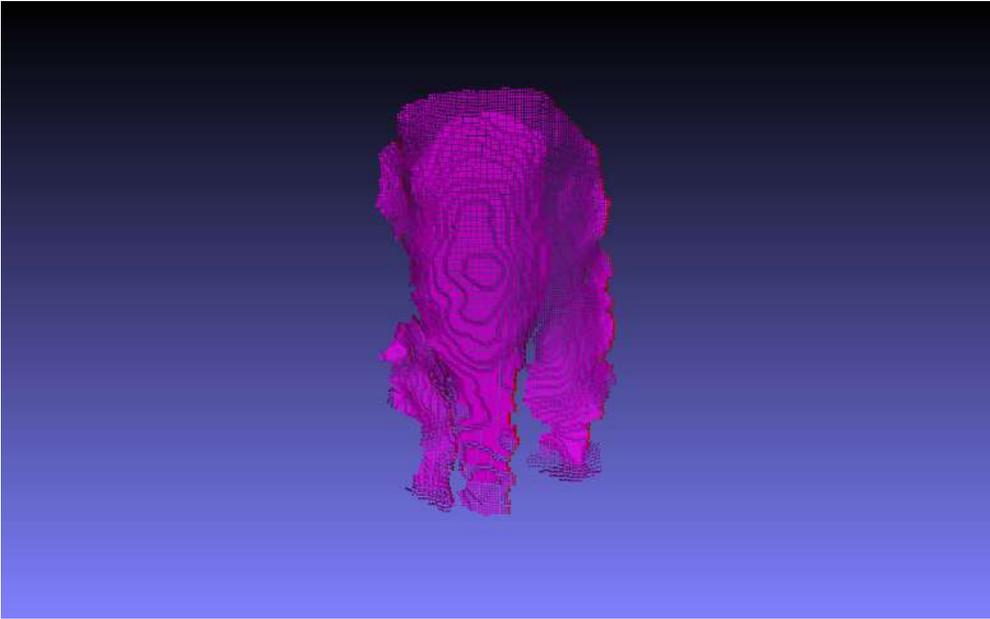


Figura 5.3: Muestra de una cara de la captura del objeto preprocesada.

- **Procesamiento de datos:** Para la reconstrucción precisa de las nubes de puntos, se empleó el método ICP implementado en el software MeshLab [52]. Este método es esencial para alinear y fusionar las nubes de puntos de manera exacta, lo que resulta en una reconstrucción detallada y precisa del objeto o escena capturados. El proceso se realizó por capas, es decir, se agregaban dos nubes de puntos, se realizaba la unión y se guardaba como una sola nube de puntos, esto se repitió hasta completar el proceso, proceso mostrado en la Figura 5.4.
- **Algoritmo de referencia:** El siguiente paso consistió en implementar los algoritmos de referencia seleccionados. En esta etapa, se emplearon los métodos ICP, Fast-ICP [53], Robust-ICP [54] y DCP [5] con el fin de realizar una comparativa exhaustiva con los resultados obtenidos tras procesar los datos utilizando el algoritmo basado en CNN previamente seleccionado. Los resultados de dicha comparación se presentan en la Tabla 5.1.
- **Pruebas:** Se llevaron a cabo pruebas exhaustivas del algoritmo de referencia con el fin de garantizar que cumpliera con los estándares previamente establecidos. Durante este proceso, se revisó minuciosamente la selección y la implementación del algoritmo,

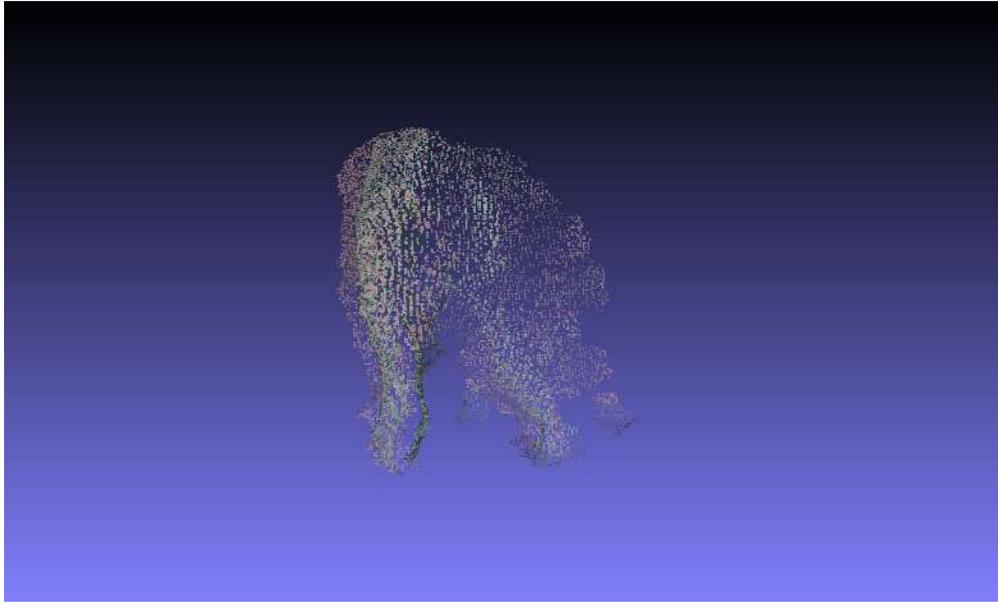


Figura 5.4: Unión de dos nubes de puntos del objeto.

realizando los ajustes necesarios para asegurar que los datos obtenidos fueran de la más alta calidad y utilidad para el desarrollo del proyecto.

- **Selección de arquitectura:** Una vez que las pruebas con el algoritmo de referencia arrojen resultados satisfactorios, el siguiente paso será la selección e implementación del modelo adecuado para el análisis de los datos. Este modelo fue entrenado meticulosamente con el objetivo de alcanzar la mayor precisión posible. Este enfoque iterativo busca optimizar continuamente el proceso, asegurando que cada etapa contribuya a la creación de un modelo robusto y eficaz. La meta final es desarrollar un sistema que no solo sea preciso, sino también confiable y eficiente en su aplicación.
- **Reconstrucción:** Para la reconstrucción precisa de las nubes de puntos, empleamos el algoritmo Transformer. Este método fue esencial para alinear y fusionar las nubes de puntos de manera exacta, resultando en una reconstrucción detallada y precisa del objeto capturado.
- **Pruebas y evaluación:** Una vez obtenidos los resultados esperados con el algoritmo, se realizaron pruebas exhaustivas al sistema para analizar en profundidad los resultados obtenidos; durante este proceso se prestó especial atención a la eficacia, precisión y

Tabla 5.1: Comparativa de ICP con diferentes porcentajes de submuestreo.

Porcentaje de submuestreo	Número de muestras	Densidad de puntos	Tiempo de procesamiento por capa (s)
50 %	10	88,510	0.234
	18	164,203	0.352
	36	325,522	0.437
60 %	10	106,211	0.247
	18	191,044	0.377
	36	390,626	0.482
70 %	10	123,913	0.259
	18	229,884	0.398
	36	455,730	0.507
80 %	10	141,617	0.304
	18	262,728	0.412
	36	520,841	0.521
90 %	10	158,015	0.322
	18	294,265	0.422
	36	584,642	0.535
100 %	10	177,026	0.364
	18	328,419	0.461
	36	651,067	0.571

rendimiento del sistema comparado con los algoritmos de referencia, con el objetivo de identificar las áreas de mejora y optimización

- Resultados:** El resultado final es un modelo tridimensional que refleja con precisión tanto la forma como las dimensiones del objeto capturado. Este modelo, generado mediante la aplicación de técnicas de registro de nubes de puntos, demuestra la capacidad del proceso para preservar los detalles geométricos esenciales del objeto. En la Figura 5.5, se presenta una representación visual de dicho modelo, donde es posible observar la fidelidad del registro en comparación con el objeto original, confirmando la eficacia de los métodos utilizados.

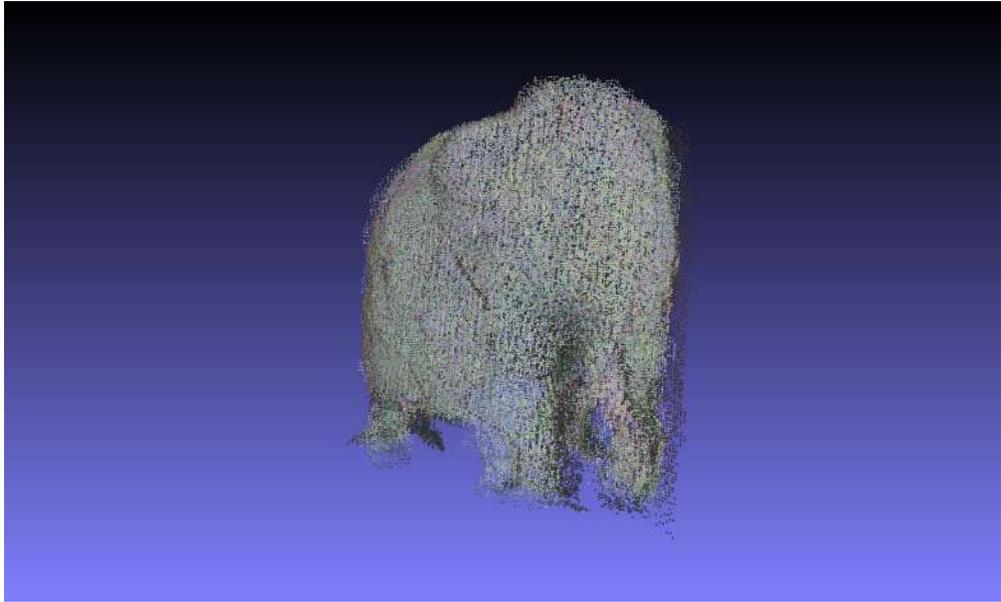


Figura 5.5: Reconstrucción mediante ICP.

Una vez aplicado el método iterativo, se obtiene un resultado similar al que se presenta en la Figura 5.6, donde se ilustra el alineamiento progresivo de las nubes de puntos, evidenciando la exactitud alcanzada en el registro tridimensional.

En este análisis, se puede observar que, aunque el método demuestra ser eficiente en términos de velocidad, no alcanza el nivel óptimo de exactitud disponible con otras técnicas actuales. Esto se debe a que el algoritmo, en cada iteración, intenta encontrar la mejor coincidencia entre los puntos. Sin embargo, en algunos casos, puede identificar coincidencias erróneas donde no las hay, lo que resulta en una nube de puntos donde ciertas caras del objeto completo no están completamente alineadas con el conjunto total de puntos. Este comportamiento se aprecia en la Figura 5.7, donde los círculos rojos resaltan las áreas en las que el algoritmo no logra converger de manera adecuada.

En el contexto de los métodos iterativos para la reconstrucción de nubes de puntos, los resultados obtenidos en este proyecto se enriquecen con un estudio previo que publiqué en la revista *La Mecatrónica en México*. En ese artículo, se realizó una comparación exhaustiva entre los métodos ICP, Fast-ICP y Robust-ICP, evaluando su desempeño en términos de eficiencia y exactitud bajo distintos escenarios de reconstrucción 3D. Los hallazgos de dicho estudio refuerzan las conclusiones del presente trabajo, destacando que el método ICP sigue

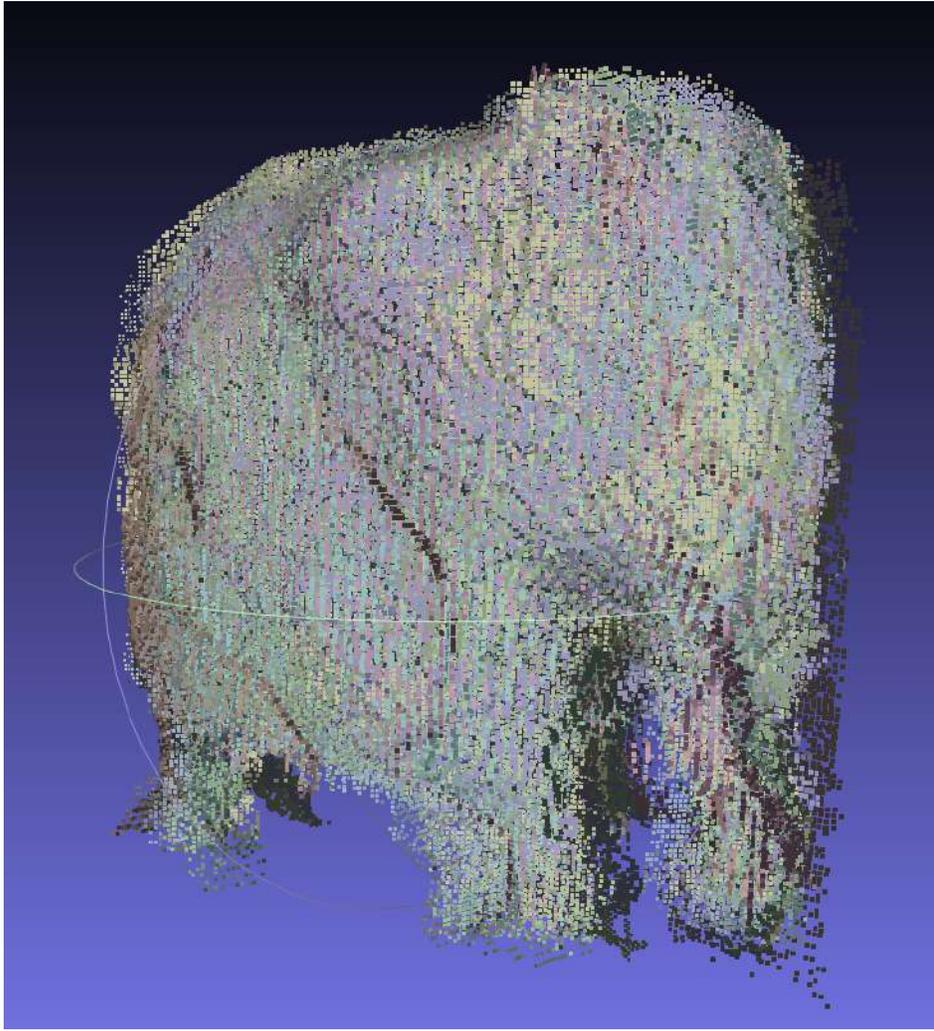


Figura 5.6: Objeto reconstruido con ICP.

siendo una opción robusta y confiable frente a datos ruidosos, siempre que se implementen técnicas de preprocesamiento adecuadas que optimicen la calidad de las nubes de puntos antes de su alineación.

Como se muestra en el artículo previamente publicado, el método Fast-ICP ofrece ventajas significativas en términos de velocidad de convergencia, pero el método ICP tradicional demostró una mayor estabilidad en la precisión de la reconstrucción cuando se trabaja con nubes de puntos de mayor densidad, como se observa en los resultados presentados en la Figura 5.5 de esta tesis. Estos hallazgos coinciden con los obtenidos en el presente estudio, donde el método ICP logra una menor tasa de error RMSE en situaciones similares.

El análisis comparativo presentado en mi artículo permitió identificar las fortalezas y

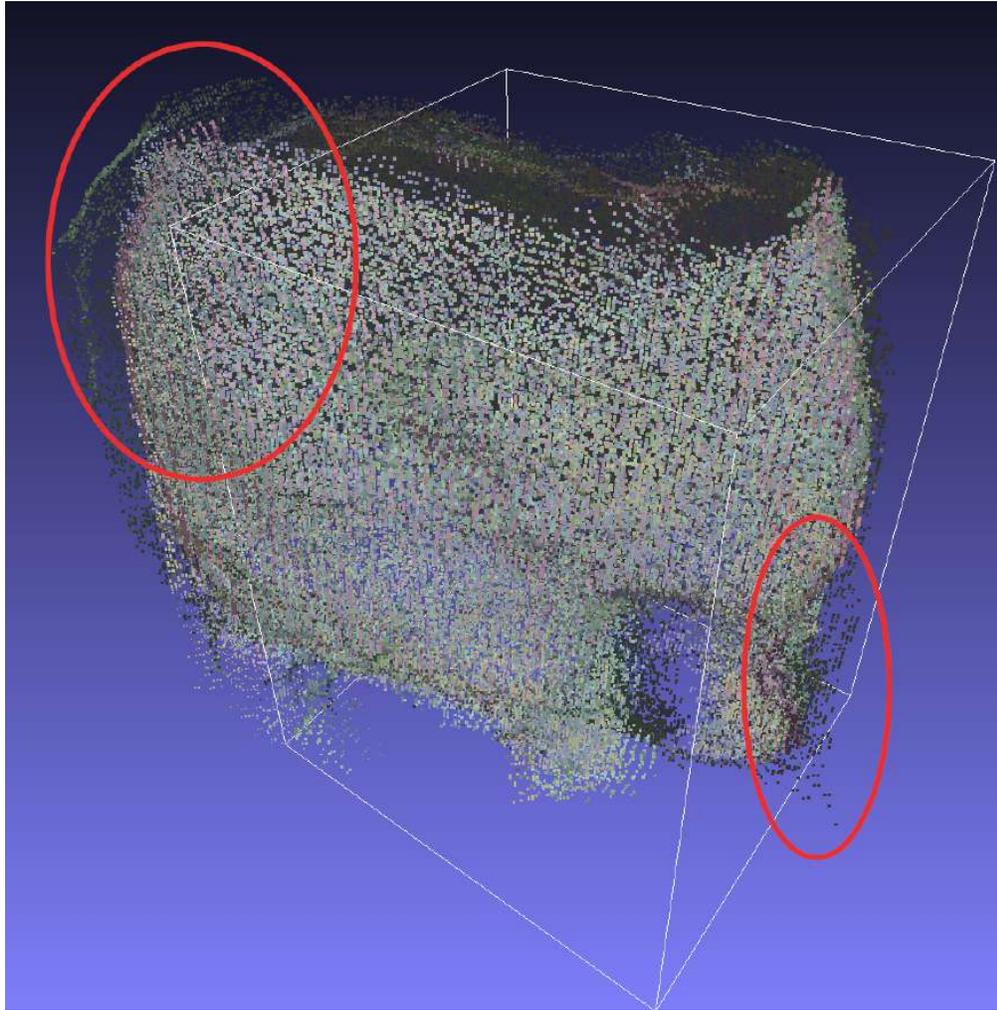


Figura 5.7: Objeto reconstruido con ICP.

debilidades de cada método iterativo, lo cual influyó directamente en la elección del método más adecuado para el procesamiento de las nubes de puntos en este proyecto. Esta experiencia contribuyó significativamente al enfoque adoptado en este trabajo, permitiendo optimizar el rendimiento del algoritmo en el conjunto de datos específico utilizado en esta investigación.

Para más detalles sobre la comparación entre los métodos iterativos, se puede consultar el trabajo publicado en “Evaluación de métodos iterativos de registro de nubes de puntos y su uso en reconstrucción 3D de objetos.”, *La Mecatrónica en México*, vol. 13, no. 2, pp. 40–58, May 2024. Online. Disponible en: www.mecamex.net/revistas/LMEM, ISSN 2448-7031 [55].

La reconstrucción 3D utilizando sensores RGB-D y el algoritmo Transformer es una téc-

nica avanzada en el campo de la visión por computadora y la robótica [56]. Con esta metodología, se buscó alcanzar e incluso superar el nivel de los estándares actuales en el campo, obteniendo un modelo altamente competitivo y a la vanguardia de la tecnología actual.

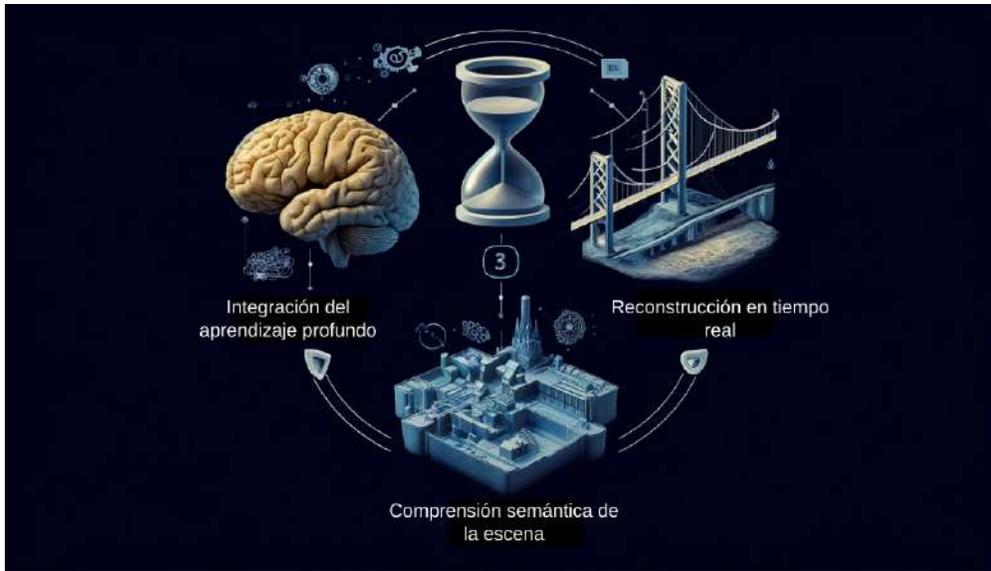


Figura 5.8: Proceso de reconstrucción 3D.

5.2. Materiales

A continuación, se detalla cada uno de los componentes y cómo se integran en el proceso de reconstrucción 3D para este proyecto.

5.2.1. Equipo de cómputo

La Tabla 5.2 hace referencia al hardware empleado para la realización de este proyecto.

5.2.2. Software

La Tabla 5.3 muestra el software empleado así como una breve descripción del uso que se le dio para garantizar el éxito de este proyecto.

Tabla 5.2: Hardware empleado y breve descripción.

Dispositivo	Descripción
Sistema Operativo	Linux Mint 21-1 Cinnamon
CPU	Intel i7-9750H @ 2.60 GHz, 6 cores
GPU	GeForce GTX 1660Ti 6 GB VRAM GDDR6
RAM	16 GB DDR4 @ 2666 MHz
Almacenamiento	1 Tb HDD + 1 Tb SSD
Sensor	Intel RealSense D415

Tabla 5.3: Software empleado y breve descripción.

Software	Descripción
MeshLab v2023.12	Utilizado para realizar el pre-procesamiento de las nubes de puntos, así como el método ICP
CloudCompare v2.13.0	Empleado para realizar el submuestreo de las capturas obtenidas
Intel RealSense Viewer v2.54.2	Utilizado para realizar las capturas de las nubes de puntos junto con el sensor descrito en la Tabla 2.1
Python v3.8	Empleado para la implementación del modelo GeoTransformer para reconstruir el objeto escaneado.

5.2.3. Base de datos

Para el desarrollo de este proyecto, se hizo uso de bases de datos públicas, como la propuesta por Uy et al. [57] en su ScanObjectNN DataSet, lo que nos permitirá obtener la información necesaria sin necesidad de someter a seres vivos a pruebas.

ModelNet-40 [58] es una conocida base de datos utilizada en el campo de la visión por computadora, específicamente en tareas de reconocimiento y clasificación de objetos en entornos tridimensionales. La base de datos empleada para el entrenamiento y pruebas, se conforma de 12,311 modelos CAD 3D pre-alineados de 40 categorías de objetos comunes como sillas, mesas, escritorios, lámparas, automóviles, etc., algunos modelos mostrados en la Figura 5.9. Cada modelo en ModelNet-40 se proporciona en formato de malla 3D (OFF), estas mallas 3D representan la forma y la geometría del objeto en cuestión. Los modelos pueden

variar en tamaño, complejidad, resolución y densidad de la malla.



Figura 5.9: 3D shapenets: A deep representation for volumetric shapes. (Z. Wu et al., 2015).

Siguiendo la metodología propuesta en [59] y [60], utilizamos los datos preprocesados de [61], los cuales muestrean uniformemente 2048 puntos de la superficie de cada modelo CAD. Para asegurar consistencia en la representación de los objetos, primero normalizamos cada modelo CAD a una unidad esférica. Posteriormente, aplicamos la misma estrategia que en [60] para generar tanto la nube de puntos fuente como la nube objetivo. Este proceso incluye el muestreo y desplazamiento de un espacio con una dirección aleatoria, manteniendo una proporción de puntos definida por el parámetro p .

La nube de puntos fuente se transforma aplicando rotaciones aleatorias en el rango de $[0, r]$ y desplazamientos aleatorios en el intervalo de $[-0,5, 0,5]$. Además, a ambas nubes de puntos se les añade ruido gaussiano con una media de 0 y una desviación estándar de 0,01,

truncado dentro del rango de $[-0,5, 0,5]$. Para la evaluación final, se seleccionan al menos 717 puntos de manera aleatoria e independiente de cada nube de puntos, generando así los pares finales de nubes a ser comparados.

Para evaluar el desempeño de nuestro método, consideramos dos escenarios de solapamiento: ModelNet con $p = 0,7$ y ModelLoNet con $p = 0,5$. Asimismo, probamos el sistema bajo dos configuraciones de rotación: una rotación pequeña con $r = 45$ y otra rotación grande con $r = 180$. Estas pruebas permiten analizar la robustez del método bajo diferentes condiciones de solapamiento y transformaciones geométricas.

Las capturas obtenidas por el sensor RealSense D415 se generaron en formato PLY. Sin embargo, el algoritmo de procesamiento empleado en este proyecto requiere que las nubes de puntos estén en formato OFF, un formato más adecuado para la gestión de geometrías tridimensionales. El formato OFF permite representar tanto los vértices como las conexiones entre ellos, lo que facilita el procesamiento de la topología del objeto. Debido a esta diferencia de formato, fue necesario convertir las nubes de puntos de PLY a OFF.

Para llevar a cabo dicha conversión, se utilizó la herramienta en línea “Image to STL” [62], la cual garantiza una conversión precisa entre formatos 3D, conservando tanto la información geométrica como los atributos visuales del objeto escaneado. Esta herramienta resultó esencial para asegurar que las nubes de puntos capturadas pudieran ser procesadas sin problemas por el algoritmo ICP, minimizando cualquier pérdida de datos relevantes.

El uso del formato OFF también proporcionó una mayor flexibilidad en las etapas posteriores de procesamiento, como la alineación y el registro de las nubes de puntos. Esto fue clave para obtener una representación tridimensional precisa del objeto escaneado, integrando correctamente los diferentes puntos de vista capturados por el sensor y facilitando una reconstrucción fiel y detallada.

5.2.4. Algoritmo.

Los Transformers, originalmente popularizada en el campo del procesamiento del lenguaje natural, en el ámbito de la visión por computadora, estos modelos han demostrado ser eficaces para capturar relaciones complejas en datos 3D y se han utilizado en una variedad de tareas de reconstrucción de objetos ofreciendo una nueva y poderosa herramienta, permitien-

do capturar relaciones complejas en datos tridimensionales y abriendo nuevas posibilidades en campos como la robótica, la realidad aumentada, la navegación autónoma y más.

Los Transformers se han utilizado en una variedad de aplicaciones prácticas, incluida la alineación y fusión de nubes de puntos capturadas desde múltiples cámaras o sensores, la generación de modelos 3D a partir de imágenes 2D y la reconstrucción de escenas tridimensionales a partir de datos de sensores, como LiDAR o RGB-D.

Teniendo dos nubes de puntos $\mathcal{P} = \{p_i \in \mathbb{R}^3 \mid i = 1, \dots, N\}$ y $\mathcal{Q} = \{q_i \in \mathbb{R}^3 \mid i = 1, \dots, M\}$, nuestro objetivo es estimar una transformación rígida $\mathbf{T} = \{\mathbf{R}, \mathbf{t}\}$ la cual alinea las dos nubes de puntos con una rotación 3D $\mathbf{R} \in \mathcal{SO}(3)$ y una traslación 3D $\mathbf{t} \in \mathbb{R}^3$. Esta transformación puede ser resuelta con la Ecuación 5.1.

$$\min_{\mathbf{R}, \mathbf{t}} \sum_{(\mathbf{p}_{x_i}, \mathbf{q}_{y_i}^*) \in \mathcal{C}^*} \|\mathbf{R} \cdot \mathbf{p}_{x_i}^* + \mathbf{t} - \mathbf{q}_{y_i}^*\|_2^2. \quad (5.1)$$

Dado que \mathcal{C}^* representa el conjunto de correspondencias de ground-truth entre \mathcal{P} y \mathcal{Q} ; y debido a que \mathcal{C}^* es desconocido, primero necesitamos establecer las correspondencias de puntos entre nuestras dos nubes de puntos y luego estimar la transformación de alineación.

Muestreo de Superpuntos y Extracción de Características

Utilizamos un eje KPConv-FPN [63], [64] para extraer las características multinivel de nuestras nubes de puntos. Uno de los beneficios adicionales del aprendizaje de características puntuales es la reducción del tamaño de la muestra de puntos; trabajamos con una muestra reducida de puntos ya que el registro de nubes de puntos puede establecerse utilizando las correspondencias de un subconjunto mucho más pequeño, esto, debido a que las nubes de puntos originales suelen ser excesivamente densas, lo que hace que las correspondencias entre puntos sean redundantes y, en ocasiones, demasiado cercanas entre sí para ser útiles.

Los puntos correspondientes a la resolución más gruesa, descritos por $\hat{\mathcal{P}}$ y $\hat{\mathcal{Q}}$, se consideran como superpuntos a emparejar. Las características aprendidas asociadas a estos superpuntos se describen como $\hat{\mathbf{F}}^{\mathcal{P}} \in \mathbb{R}^{|\hat{\mathcal{P}}| \times \hat{d}}$ y $\hat{\mathbf{F}}^{\mathcal{Q}} \in \mathbb{R}^{|\hat{\mathcal{Q}}| \times \hat{d}}$. Las correspondencias de puntos densos son calculadas a la mitad de la resolución original, por ejemplo, los puntos submuestreados de primer nivel se describen como $\tilde{\mathcal{P}}$ y $\tilde{\mathcal{Q}}$; sus características aprendidas se representan como

$$\tilde{\mathbf{F}}^{\mathcal{P}} \in \mathbb{R}^{|\tilde{\mathcal{P}}| \times \tilde{d}} \text{ y } \tilde{\mathbf{F}}^{\mathcal{Q}} \in \mathbb{R}^{|\tilde{\mathcal{Q}}| \times \tilde{d}}.$$

Para cada superpunto, se construyen parches locales de puntos a su alrededor utilizando una estrategia de agrupación punto-a-nodo [65], [66], lo que da lugar a una descomposición de Voronoi, como se muestra en la Figura 5.10. En esta representación, se puede observar que cada punto $\tilde{\mathcal{P}}$ y $\tilde{\mathcal{Q}}$, junto con sus características $\tilde{\mathbf{F}}^{\mathcal{P}}$ y $\tilde{\mathbf{F}}^{\mathcal{Q}}$, son asignados al superpunto más cercano en el espacio geométrico, de acuerdo con la Ecuación 5.2.

$$\mathcal{G}_i^{\mathcal{P}} = \{\tilde{\mathbf{p}} \in \tilde{\mathcal{P}} \mid i = \operatorname{argmin}_j (\|\tilde{\mathbf{p}} - \hat{\mathbf{p}}_j\|), \hat{\mathbf{p}}_j \in \hat{\mathcal{P}}\} \quad (5.2)$$

En la Figura 5.10 se muestra de manera gráfica el proceso de descomposición descrito previamente. A la izquierda, se observa la nube de puntos (en azul) y la muestra de superpuntos (en rojo). A la derecha, se presenta una asignación por color de los puntos, según el superpunto al que fueron asignados.

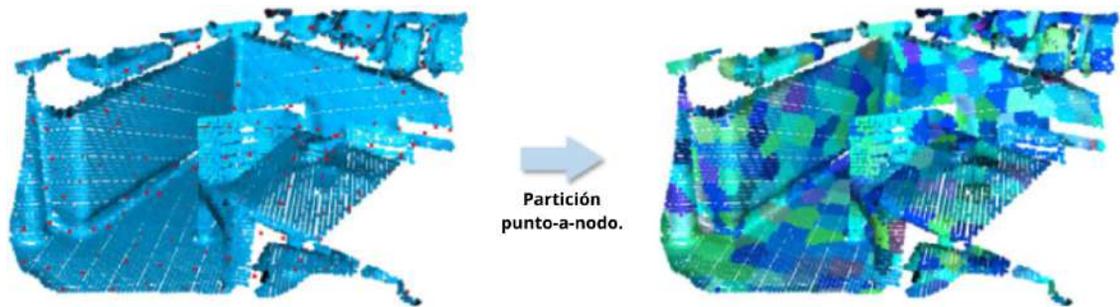


Figura 5.10: Estrategia de agrupación punto-a-nodo.

La matriz de características asociada a los puntos en $\mathcal{G}_i^{\mathcal{P}}$ se describe como $\mathbf{F}^{\mathcal{P}}_i \subset \tilde{\mathbf{F}}^{\mathcal{P}}$, y los parches vacíos se eliminan. De manera similar, los parches para $\mathcal{G}_i^{\mathcal{Q}}$ y las matrices de características $\mathbf{F}^{\mathcal{Q}}_i$ para la nube de puntos \mathcal{Q} se calculan y describen de la misma forma.

Modulo de Coincidencia de Superpuntos

Debido a que los métodos actuales generalmente alimentan al Transformer con características de alto nivel extraídas de las nubes de puntos [59], [65], [5], sin codificar explícitamente la estructura geométrica, las características aprendidas tienden a ser menos discriminatorias. Esto genera una mayor ambigüedad en las coincidencias y puede resultar en coincidencias atípicas, especialmente en casos de baja superposición.

En este contexto, Qin et al. [56] propusieron el Transformer geométrico, que no solo codifica características de alto nivel, sino que también captura explícitamente las estructuras geométricas intra-nube y la consistencia geométrica inter-nube. Este Transformer se compone de un módulo de auto-atención geométrica, utilizado para el aprendizaje de características dentro de las nubes de puntos, y un módulo de atención basado en características para modelar la coherencia entre las nubes de puntos. Estos dos módulos se intercalan N_t veces para extraer las características híbridas de \hat{H}^P y \hat{H}^Q , como se muestra en la Figura 5.11, lo que permite una coincidencia de superpuntos confiable.

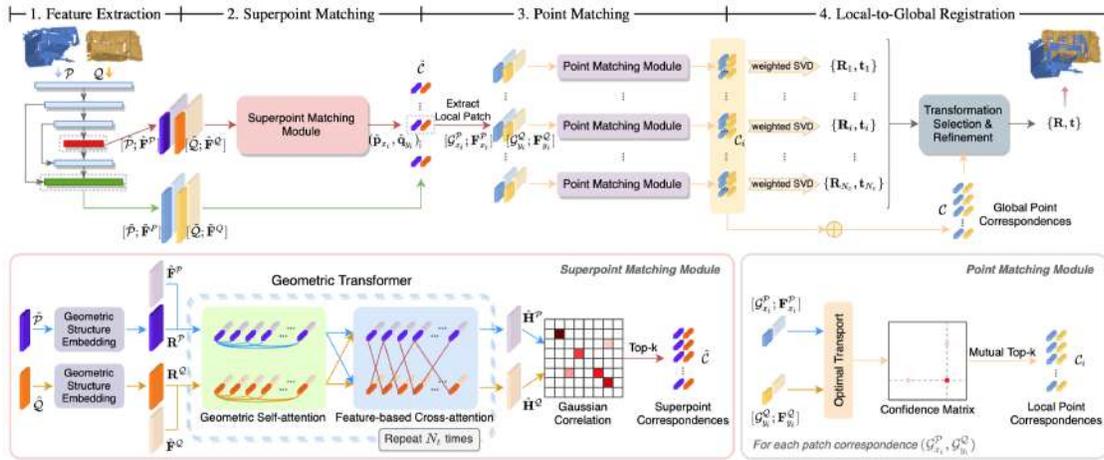


Figura 5.11: Transformer geométrico para registro de nubes de puntos.

Auto-atención Geométrica

Qin et al. [56] diseñaron el módulo de auto-atención geométrica, que aprende las relaciones globales tanto de las características como de los espacios geométricos entre los superpuntos de cada nube de puntos. A continuación, se describe el cálculo tanto para \hat{P} como para \hat{Q} . Dada nuestra matriz de entrada $\mathbf{X} \in \mathbb{R}^{|\hat{P}| \times d_t}$, la matriz de salida $\mathbf{Z} \in \mathbb{R}^{|\hat{P}| \times d_t}$ es la suma ponderada de todas las características de entrada, y se calcula mediante la Ecuación 5.3:

$$\mathbf{z}_i = \sum_{j=i}^{|\hat{P}|} a_{i,j} (\mathbf{x}_j \mathbf{W}^V) \quad (5.3)$$

el coeficiente de peso $a_{i,j}$ se obtiene aplicando una función softmax a nivel de filas sobre el puntaje de atención $e_{i,j}$. El valor de $e_{i,j}$ se calcula según la Ecuación 5.4, que determina la intensidad de la correlación entre el punto i y el punto j en el espacio de atención.

$$e_{i,j} = \frac{(\mathbf{x}_i \mathbf{W}^Q)(\mathbf{x}_j \mathbf{W}^K + r_{i,j} \mathbf{W}^R)^T}{\sqrt{d_t}} \quad (5.4)$$

donde $r_{i,j}$ representa la integración de las estructuras geométricas, como se describirá a continuación. Las matrices \mathbf{W}^Q , \mathbf{W}^K , \mathbf{W}^V y $\mathbf{W}^R \in \mathbb{R}^{d_t \times d_t}$ son las proyecciones para Q (query), K (key), V (value) y R (integración de estructuras geométricas), respectivamente. La Figura 5.12 ilustra la estructura y el proceso de cálculo del módulo de autoatención geométrica.

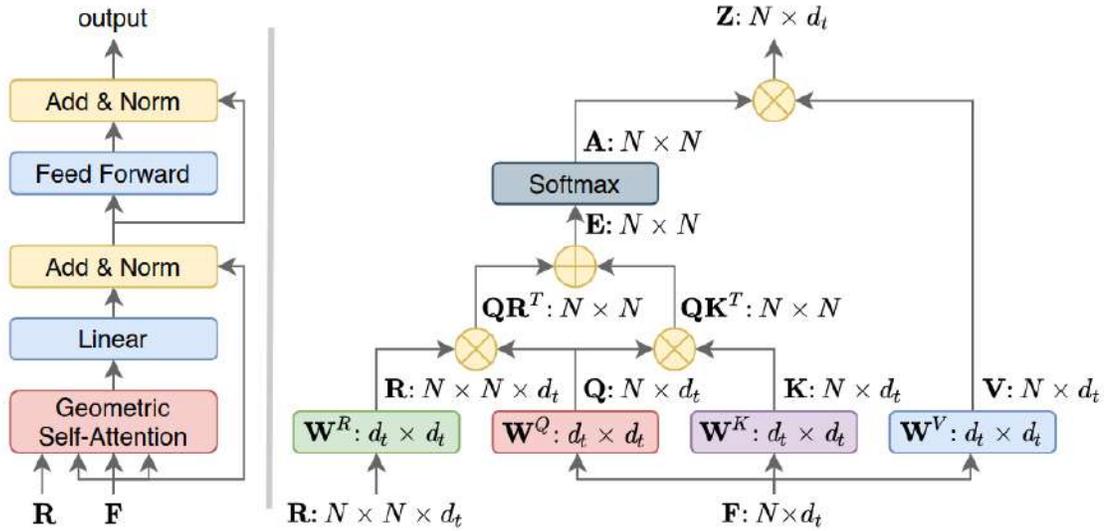


Figura 5.12: Módulo de auto-atención geométrica.

La integración de estructuras geométricas se diseñó para codificar la estructura geométrica invariante de las transformaciones en los superpuntos. La idea central es medir las distancias y ángulos calculados con los superpuntos, los cuales son consistentes a través de diferentes nubes de puntos de la misma escena; es decir, dados dos superpuntos $\hat{\mathbf{p}}_i, \hat{\mathbf{p}}_j \in \hat{\mathbf{P}}$, la integración de sus estructuras geométricas consiste en:

- **Integración por parejas** dadas las distancias $\rho_{i,j} = \|\hat{\mathbf{p}}_i - \hat{\mathbf{p}}_j\|$ entre $\hat{\mathbf{p}}_i$ y $\hat{\mathbf{p}}_j$ calculada con la Ecuación 5.5.

$$\begin{cases} \mathbf{r}_{i,j,2k}^D = \sin\left(\frac{d_{i,j}/\sigma_d}{10000^{2k/d_t}}\right) \\ \mathbf{r}_{i,j,2k+1}^D = \cos\left(\frac{d_{i,j}/\sigma_d}{10000^{2k/d_t}}\right) \end{cases} \quad (5.5)$$

donde σ_d es un hiperparametro de temperatura empleado para ajustar la sensibilidad de las variaciones de distancia.

- **Integración angular triple** primero seleccionamos los K vecinos cercanos \mathcal{K}_i de \hat{p}_i . Para cada $\hat{p}_x \in \mathcal{K}_i$, calculamos el ángulo $\alpha_{i,j}^x =$ se obtiene mediante la Ecuación 5.6

$$\begin{cases} \mathbf{r}_{i,j,x,2k}^A = \sin\left(\frac{\alpha_{i,j}^x/\sigma_a}{10000^{2l/d_t}}\right) \\ \mathbf{r}_{i,j,x,2k+1}^A = \cos\left(\frac{\alpha_{i,j}^x/\sigma_a}{10000^{2l/d_t}}\right) \end{cases} \quad (5.6)$$

donde σ_a controla la sensibilidad de las variaciones angulares.

Por último, la integración de las estructuras geométricas, $\mathbf{r}_{i,j}$ se calcula con la Ecuación 5.7

$$\mathbf{r}_{i,j} = \mathbf{r}_i^D \mathbf{W}^D + \max_x \{ \mathbf{r}_{i,j,x}^A \mathbf{W}^A \} \quad (5.7)$$

donde $\mathbf{W}^D, \mathbf{W}^A \in \mathbb{R}^{d_t \times d_t}$ son las matrices de proyección correspondientes a ambos tipos de integración. Se utiliza max pooling para mejorar la robustez frente a las variaciones en los vecinos cercanos, debido a la autoexclusión. La Figura 5.13 ilustra el cálculo realizado para la integración de estructuras geométricas.

Función de pérdida

Las correspondencias de puntos reales son relativamente limitadas, ya que solo están disponibles para las nubes de puntos obtenidas a partir de un muestreo reducido. Para abordar esto, utilizamos una función de pérdida basada en la log-verosimilitud negativa en la matriz de asignación [67], específicamente en la matriz de asignación $\bar{\mathbf{Z}}_i$ para cada correspondencia de superpuntos. Durante el entrenamiento, se seleccionan aleatoriamente N_g correspondencias de superpuntos ground-truth $\{\hat{\mathcal{C}}_i\}$ en lugar de emplear las predichas. Para cada $\hat{\mathcal{C}}_i$, se

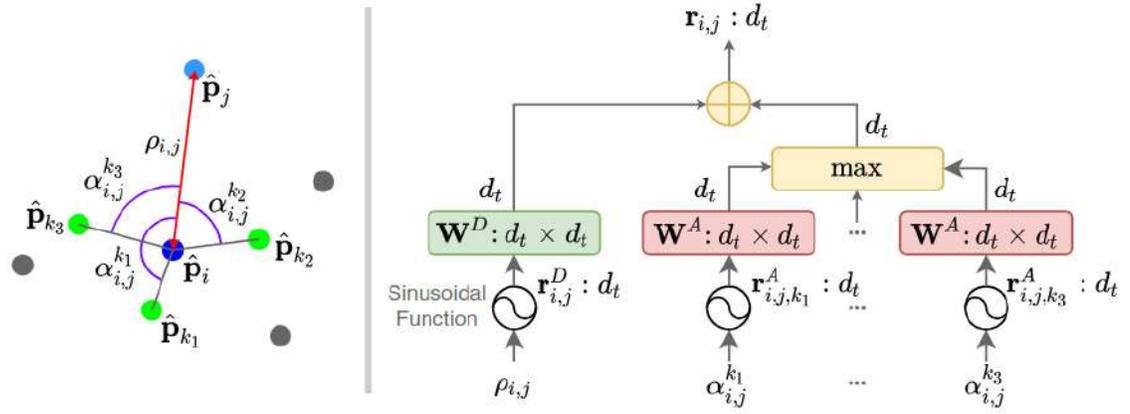


Figura 5.13: Integración de estructuras geométricas.

extrae un conjunto de correspondencias ground-truth \mathcal{M}_i utilizando un radio de coincidencia τ . Los puntos que no coinciden en los parches se denotan como \mathcal{I}_i y \mathcal{J}_i . Las pérdidas de coincidencia de puntos individuales para $\hat{\mathcal{C}}_i^*$ se calculan según la ecuación 5.8.

$$\mathcal{L}_{p,i} = \sum_{(x,y) \in \mathcal{M}_i} \log \bar{z}_{x,y}^i - \sum_{x \in \mathcal{I}_i} \log \bar{z}_{x,m_i+1}^i - \sum_{y \in \mathcal{J}_i} \log \bar{z}_{n_i+1,y}^i \quad (5.8)$$

La pérdida final se determina promediando las pérdidas individuales de todas las correspondencias de superpuntos, según la siguiente expresión: $\mathcal{L}_p = \frac{1}{N_g} \sum_{i=1}^{N_g} \mathcal{L}_{p,i}$. Este promedio permite que la función de pérdida represente de manera efectiva la desviación global del modelo en relación con las correspondencias esperadas, asegurando una optimización más equilibrada a lo largo del conjunto de datos.

Entrenamiento

En esta sección, se presentan los resultados obtenidos durante el proceso de entrenamiento de la red neuronal. Para la implementación y evaluación del modelo, se utilizó la librería PyTorch [68] ejecutada en una tarjeta gráfica Nvidia GTX 1660Ti, y se empleó el optimizador Adam [69]. El rendimiento del modelo fue evaluado utilizando el error cuadrático medio (RMSE) como métrica principal, dado que esta métrica permite cuantificar la precisión del ajuste del modelo a los datos de entrenamiento y proporciona una estimación clara de la magnitud de los errores de predicción.

Para este estudio, seguimos el protocolo descrito en [59], seleccionando las primeras 20

categorías en la división oficial de entrenamiento/prueba para el entrenamiento y la validación, mientras que las otras 20 categorías se reservaron para la fase de prueba. Adicionalmente, eliminamos 8 categorías con simetría (botella, cuenco, cono, taza, maceta, lámpara, tienda de campaña y jarrón) debido a la ambigüedad de sus poses, lo que podría afectar negativamente la precisión del modelo.

Como resultado, trabajamos con un total de 4194 modelos CAD destinados al entrenamiento, 1002 para la validación y 1146 para las pruebas. Esta selección garantiza que los modelos utilizados en cada etapa del proceso sean representativos y que los resultados obtenidos puedan ser evaluados con mayor precisión y consistencia, evitando las complicaciones que presentan las categorías con simetría.

El proceso de entrenamiento se llevó a cabo a lo largo de un total de 400,000 iteraciones, con evaluaciones periódicas realizadas cada 100 iteraciones, lo que permitió monitorear la evolución del modelo a lo largo del tiempo y ajustar los parámetros de manera óptima. La tasa de aprendizaje inicial fue de 10^{-4} , con un decrecimiento exponencial aplicado cada 5 épocas para mejorar la estabilidad y convergencia del modelo durante las fases avanzadas del entrenamiento.

Para garantizar la robustez del proceso de entrenamiento, se emplearon técnicas de aumento de datos similares a las descritas en [59]. Además, durante el entrenamiento, se muestrearon de manera aleatoria $N_g = 128$ correspondencias de súper-puntos ground-truth, mientras que para la fase de prueba se utilizaron $N_c = 256$ coincidencias putativas de súper-puntos, con el fin de evaluar el rendimiento del modelo en diferentes escenarios y mejorar su capacidad general de adaptación a datos no vistos.

La Tabla 5.4 muestra la relación entre el número de iteraciones del proceso de entrenamiento y los valores correspondientes de RMSE obtenidos. Como se puede observar, el error decrece conforme aumenta el número de iteraciones, lo cual es un comportamiento esperado en este tipo de modelos, ya que con más iteraciones el modelo mejora su capacidad de generalización y precisión.

- En las primeras 30,000 iteraciones, el valor del RMSE era considerablemente alto (0.032), lo que indica que el modelo aún no había captado correctamente los patrones subyacentes en los datos de entrada.

Tabla 5.4: Resultados RMSE en función de las iteraciones de entrenamiento.

ITER	RMSE
400,000	0.017
380,000	0.018
340,000	0.018
280,000	0.019
180,000	0.022
30,000	0.032

- A medida que aumentaban las iteraciones de entrenamiento, el RMSE se fue reduciendo progresivamente, alcanzando un valor de 0.018 alrededor de las 340,000 y 380,000 iteraciones, lo que demuestra que el modelo estaba ajustándose cada vez más a los datos.
- Finalmente, tras 400,000 iteraciones, se obtuvo el mejor valor de RMSE, 0.017, lo que indica que el modelo alcanzó una convergencia adecuada en este punto del entrenamiento.

Estos resultados indican que, si bien el modelo mejora su rendimiento con un número elevado de iteraciones, es importante encontrar un balance entre el tiempo de cómputo y el nivel de precisión obtenido. Es decir, después de un punto, el incremento en las iteraciones no necesariamente resulta en una mejora significativa en el RMSE, lo cual puede ser observado en las iteraciones superiores a 340,000, donde la mejora fue mínima.

La información obtenida a partir de este entrenamiento nos proporciona una base sólida para ajustar y optimizar los parámetros del modelo de manera efectiva en futuras investigaciones.

El entrenamiento del modelo se llevó a cabo en una configuración mostrada en las Tablas 5.2 y 5.3, lo que permitió una ejecución relativamente rápida del algoritmo. El entrenamiento de la red, a lo largo de 400,000 iteraciones, requirió aproximadamente 72 horas, lo que demuestra la importancia de un entorno computacional robusto para tareas de alta demanda de procesamiento como la reconstrucción 3D basada en redes neuronales. Aunque la precisión

del modelo ($RMSE = 0,017$) mejoró con el tiempo, a partir de la iteración 180,000 se comenzó a notar una estabilización en la mejora, lo que sugiere que aplicar técnicas como early stopping en futuras implementaciones podría reducir el tiempo de entrenamiento sin comprometer significativamente la calidad del resultado. Este análisis subraya la aplicabilidad de la red entrenada para tareas de alta precisión, como la robótica y la realidad aumentada, donde la precisión tridimensional es crítica, aunque el tiempo de procesamiento inicial sea más elevado.

5.3. Consideraciones Éticas

Es fundamental garantizar que cualquier proyecto, como las pruebas de reconstrucción 3D utilizando cámaras RGB-D, se desarrolle bajo estrictas normas éticas y de seguridad, evitando riesgos para las personas, animales, plantas o el medio ambiente. El uso de estas cámaras se enfoca únicamente en la captura y procesamiento de datos, sin intervención directa en seres vivos ni modificaciones al entorno natural. Así, se priorizan la integridad, privacidad y salud, cumpliendo con los más altos estándares éticos en todo momento.

El compromiso de no realizar experimentos en animales o seres humanos refleja una práctica ética sólida que salvaguarda el bienestar de los seres vivos. En lugar de ello, se recurrió al uso de bases de datos públicas, lo cual permitió obtener resultados de calidad sin comprometer la vida de los seres vivos.

Adicionalmente, al no involucrar fuentes de financiamiento externas a la Universidad Autónoma de Querétaro, se evita cualquier posible conflicto de interés. Esto asegura que la investigación se lleve a cabo con total independencia, garantizando que los resultados sean imparciales y no estén influenciados por intereses externos.

Capítulo 6

Resultados y discusión

Previo al desarrollo de la investigación, se publicó un artículo en la revista La Mecatrónica en México, bajo el número de registro LMEM20240401-1 [55]. En dicho artículo, se presentaron los resultados obtenidos durante el estudio comparativo de métodos de reconstrucción 3D, que se resumen en la Tabla 6.1.

Tabla 6.1: Comparativa entre métodos de reconstrucción de nubes de puntos.

Método	Tiempo	RMSE
ICP	0.33	0.053
Fast-ICP	0.12	0.031
Robust ICP	0.96	0.022
DCP	0.02	0.014
GeoTransformer	0.47	0.017

Para la unión de puntos mediante el método ICP y GeoTransformer, se utilizaron los siguientes parámetros, mostrados en la Tabla 6.2:

El proceso de registro implica iteraciones en las cuales se ajusta la rotación y traslación de una nube de puntos de manera que se alinee lo más posible con otra nube de puntos. Durante las iteraciones, se trata de minimizar la suma de las distancias euclidianas entre los puntos correspondientes de ambas nubes de puntos.

El algoritmo utilizado encontró la mejor correspondencia entre los puntos de nuestras

Tabla 6.2: Parámetros para registro de nubes de puntos.

Parámetro	Valor
Número máximo de muestras	50,000
Superposición final	100 %
Distancia objetivo	0.005 mm
RMS objetivo	1e-5
RMS real	0.0535

nubes de puntos y calculó la transformación que minimiza la distancia entre ellas, ecuación 6.1, que representa la matriz de rotación y traslación resultante de este proceso de registro.

$$Rt = \begin{pmatrix} 0,993 & -0,060 & 0,100 & 0,03 \\ 0,049 & 0,994 & 0,102 & 0,046 \\ -0,105 & -0,096 & 0,990 & -0,005 \\ 0,000 & 0,000 & 0,000 & 1,000 \end{pmatrix} \quad (6.1)$$

Los elementos de la matriz describen cómo rotar y trasladar los puntos de una nube de puntos para alinearlos con otra. Las primeras tres filas y columnas representan la rotación, mientras que la última columna define la traslación.

La Figura 6.1 muestra por separado la información de profundidad y color obtenida en estas capturas. Esta representación diferenciada resalta las distintas dimensiones de los datos capturados: la textura y el color superficial del objeto, junto con su estructura tridimensional. Estos aspectos son esenciales para lograr una comprensión completa y precisa del objeto estudiado.

Al utilizar el sensor RGB-D para la captura, se obtiene una representación detallada del objeto que combina información de color y profundidad. Esta integración de datos genera una imagen rica y tridimensional, que facilita una comprensión más completa del objeto en estudio. La Figura 6.2 presenta el resultado de esta captura, donde las tonalidades y texturas del objeto se complementan con la información de profundidad, ofreciendo una visión integral y precisa desde múltiples perspectivas.

Además, los números mostrados indican la distancia del objeto escaneado al sensor RGB-

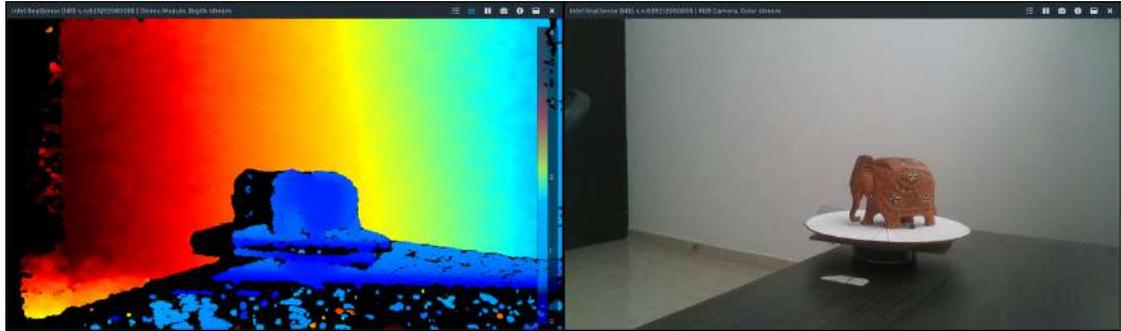


Figura 6.1: Captura de profundidad y RGB de izquierda a derecha respectivamente.

D, mientras que la estrella azul señala el centro del sensor.



Figura 6.2: Captura RGB-D del sensor RealSense de Intel.

Después de completar las capturas, se procedió a importar los archivos en formato PLY al software MeshLab. Este paso implicó la unión cuidadosa de las capturas, trabajando con cada par de caras adyacentes de forma secuencial. Este método se adoptó con el objetivo de minimizar el error en la alineación y asegurar una fusión precisa de las imágenes para formar una representación cohesiva y detallada del objeto.

Dada la sensibilidad del método ICP al ruido y a los outliers, fue esencial llevar a cabo un proceso de limpieza para cada una de las capturas, es decir, eliminar cualquier punto que no forme parte del objeto escaneado (base rotativa del objeto, cualquier tipo de ruido capturado por el sensor, etc). Este paso previo de preprocesamiento consistió en identificar y eliminar cualquier anomalía que pudiera afectar negativamente la precisión de la alineación.

Esta limpieza ayudó a asegurar que los datos utilizados en el proceso de alineación fueran lo más precisos y representativos posible del objeto real, mejorando así la calidad y fiabilidad del resultado final de la fusión de las nubes de puntos.

La alineación del objeto se efectuó identificando puntos de referencia entre dos caras adyacentes. Este procedimiento se realizó de manera iterativa, de tal manera que, tras cada paso, las dos caras en consideración se fusionaban efectivamente en una sola. Este proceso se repitió hasta conseguir la alineación completa de todas las caras, resultando en la reconstrucción integral del objeto. En la Figura 6.3 podemos observar el resultado final, después de aplicar el método ICP al objeto. En esta figura, se puede observar cómo la metodología utilizada permitió la unión precisa y detallada de todas las capturas, resultando en una representación cohesiva y tridimensional del objeto en estudio.

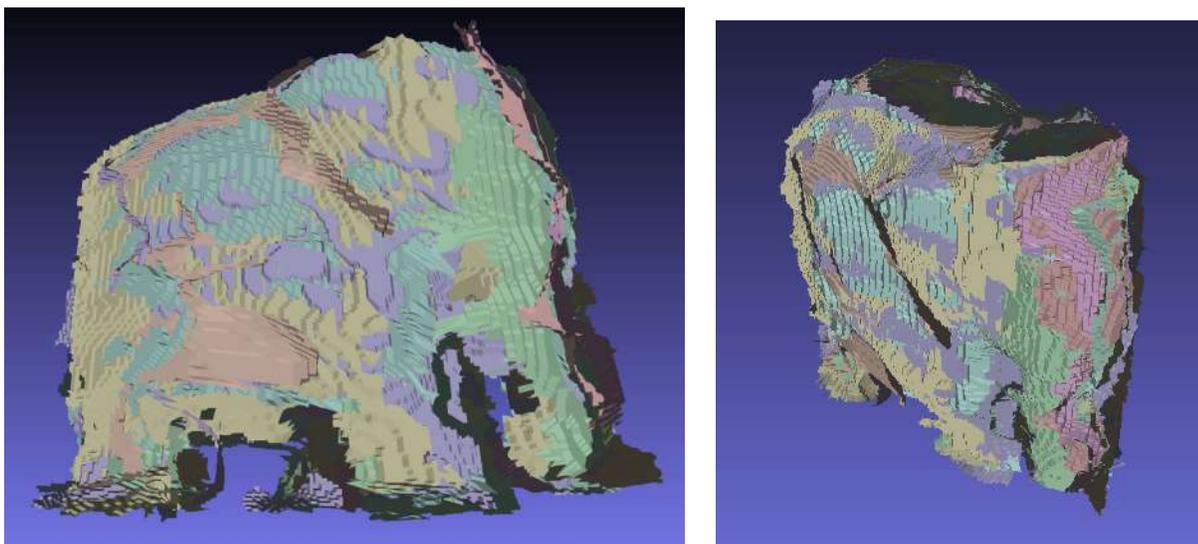


Figura 6.3: Objeto reconstruido mediante ICP sin submuestreo, 18 capturas.

Sin embargo, como se puede apreciar en la Figura 6.3, el objeto en cuestión no es completamente sólido. Esta característica se hace evidente en la visualización de las capturas, donde se observan áreas que no presentan la continuidad típica de un objeto sólido. Esta observación es crucial para entender las limitaciones y desafíos en la reconstrucción tridimensional de objetos con estas características específicas.

La Figura 6.4 muestra el objeto escaneado en su estado inicial, es decir, antes de aplicar el método ICP. En estas imágenes, se observa la nube de puntos del objeto capturada por el

sensor, que muestra una disposición desordenada y una falta de alineación entre las diferentes tomas. Se pueden apreciar variaciones en la densidad de puntos y posibles discrepancias en la orientación del objeto. Estas diferencias se destacan en la visualización en 3D, donde la forma y la estructura del objeto no están completamente definidas debido a la falta de alineación entre las capturas individuales. Esta representación inicial del objeto proporciona una base para el proceso de alineación mediante el método ICP, con el objetivo de fusionar y alinear correctamente las múltiples tomas para obtener una representación coherente y precisa del objeto escaneado.

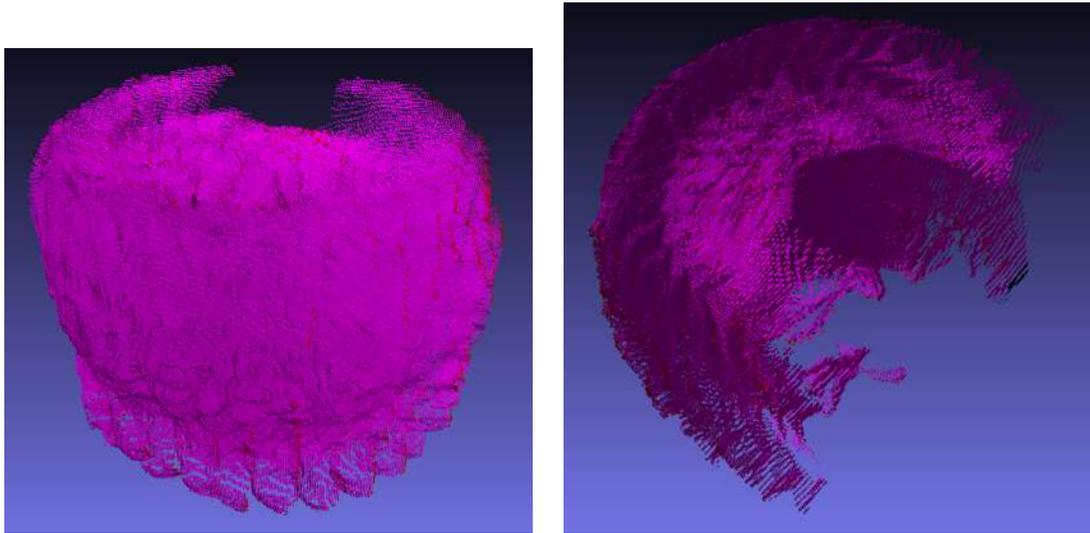


Figura 6.4: Nubes de puntos del objeto escaneado previo a la implementación de los métodos.

El submuestreo plantea el desafío de reducir la densidad de puntos, lo que acorta el tiempo y las iteraciones necesarias para que el algoritmo converja. Sin embargo, como se puede observar en la Figura 6.5, donde se compara el submuestreo al 50 %, 60 %, 70 %, 80 %, 90 % y 100 %, se evidencia que a medida que se reduce la densidad de puntos, la pieza resultante presenta espacios más grandes que requieren un tratamiento adicional para su llenado. Conforme aumenta la densidad de puntos, el método iterativo necesita más recursos para lograr una convergencia óptima.

Como se puede observar en la Figura 6.6, la densidad de puntos provoca un desfase en las partes anterior y posterior del objeto, indicado en rojo. Este desfase se debe a que el algoritmo no logra una convergencia adecuada para realizar el registro. Para solucionar este

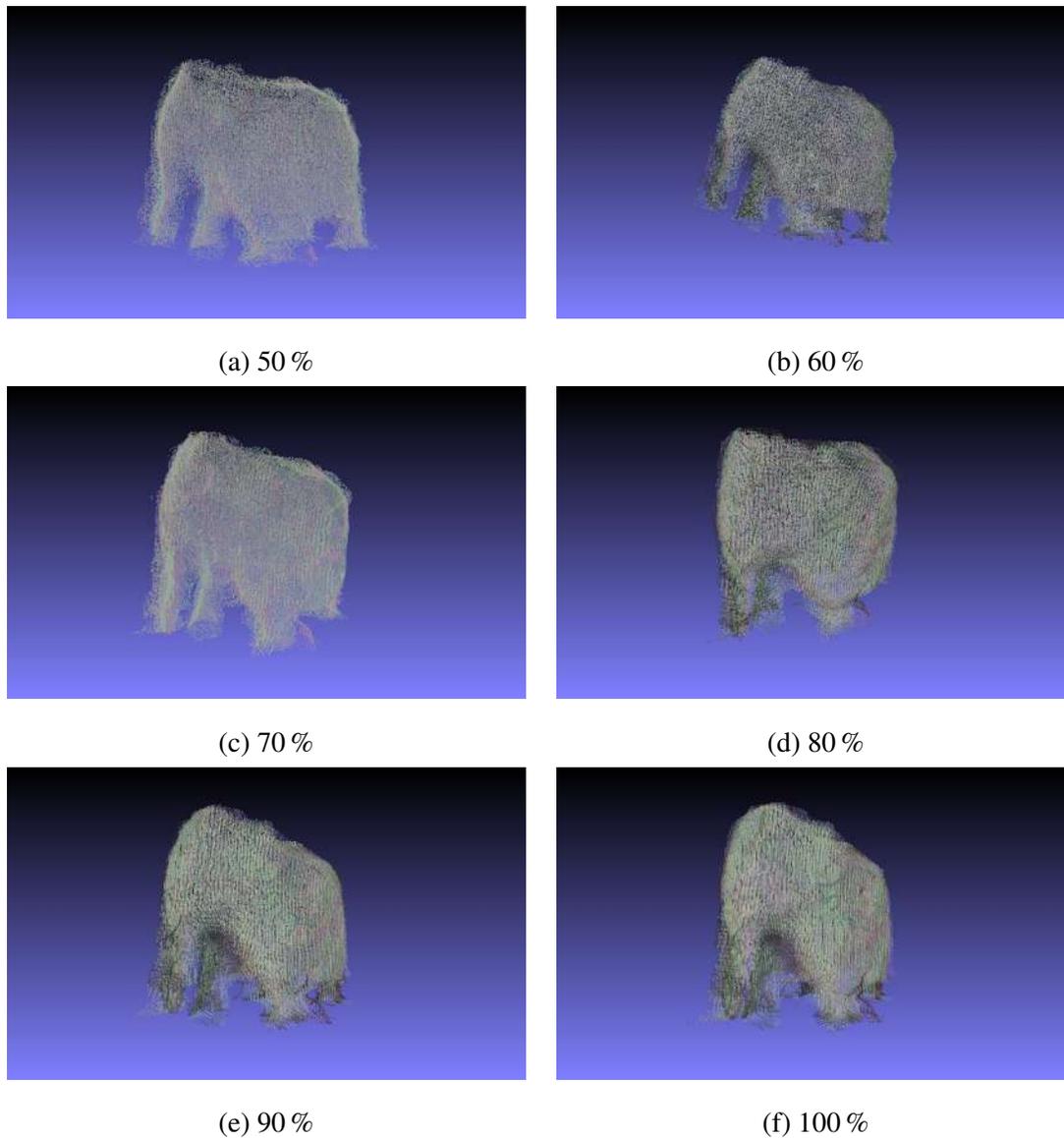


Figura 6.5: (a) Método ICP con diferentes niveles de submuestreo y 10 muestras.

inconveniente, fue necesario repetir el proceso hasta alcanzar una convergencia correcta.

La Tabla 6.3 presenta una comparativa entre las densidades al aplicar el método ICP con diferentes porcentajes de submuestreo, las muestras necesarias y el tiempo requerido para completar el proceso.

De la Tabla 6.3 se puede inferir que a medida que aumenta la densidad, el algoritmo tiende a converger más fácilmente, aunque se vuelve más susceptible al ruido y a errores; esto debido a que, con una mayor densidad de puntos, hay una mayor probabilidad de encontrar

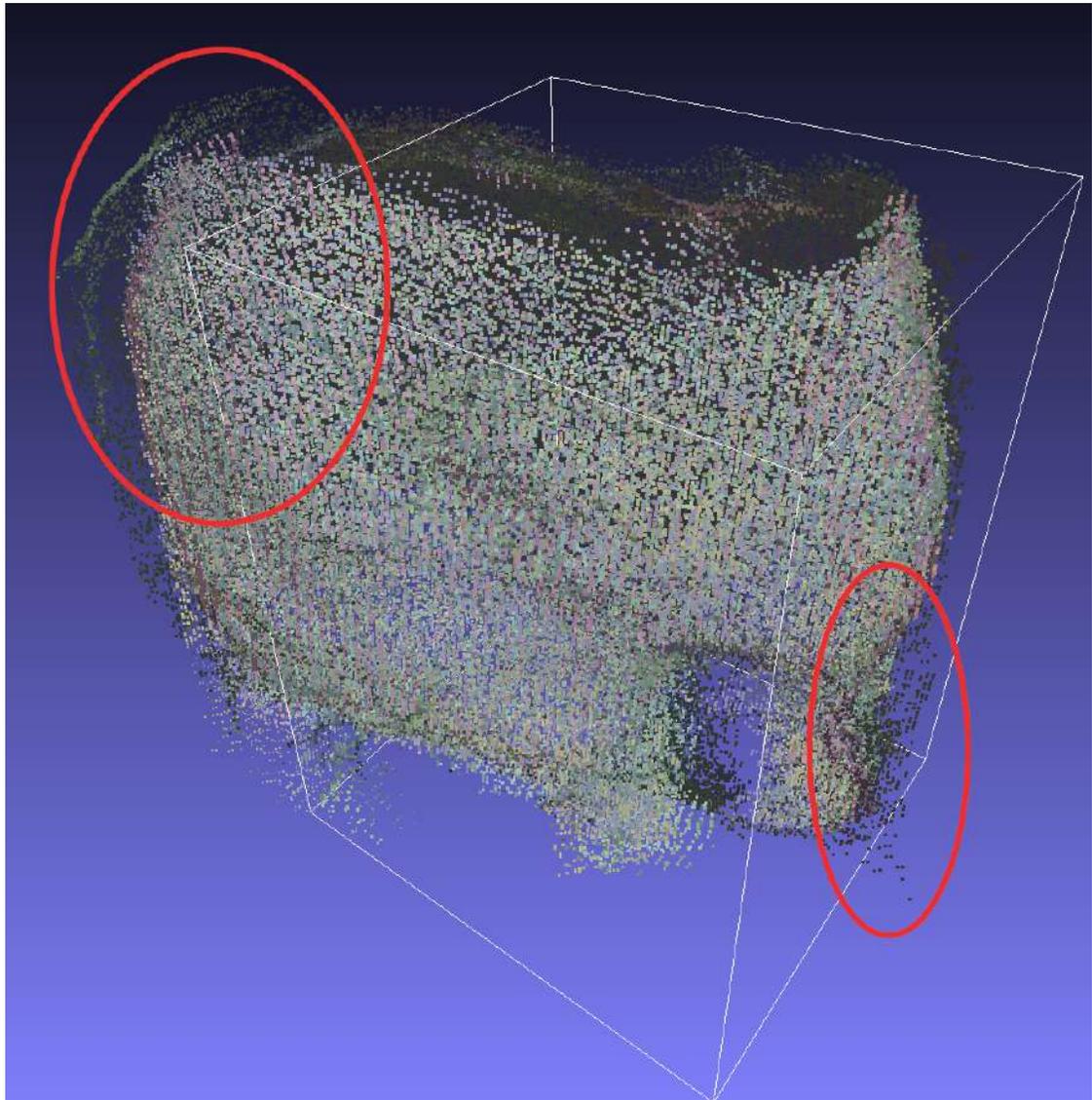


Figura 6.6: Captura de profundidad y RGB de izquierda a derecha respectivamente.

similitudes entre ellos, lo que puede resultar en un registro incorrecto.

Como se ilustra en la Figura 6.7, se puede observar un desfase en la zona marcada con un círculo rojo, donde una de las nubes de puntos no está correctamente alineada con el conjunto total.

Dado que el objeto escaneado no presenta superficies translúcidas y la base de datos utilizada es relativamente pequeña, compuesta por 36 muestras, el tiempo de ejecución del proceso no se vio afectado de manera significativa.

Además, la ausencia de superficies translúcidas y el tamaño manejable del conjunto de

Tabla 6.3: Comparativa de ICP con diferentes porcentajes de submuestreo.

Porcentaje de submuestreo	Número de muestras	Densidad de puntos	Tiempo de procesamiento por capa (s)
50 %	10	88,510	0.234
	18	164,203	0.352
	36	325,522	0.437
60 %	10	106,211	0.247
	18	191,044	0.377
	36	390,626	0.482
70 %	10	123,913	0.259
	18	229,884	0.398
	36	455,730	0.507
80 %	10	141,617	0.304
	18	262,728	0.412
	36	520,841	0.521
90 %	10	158,015	0.322
	18	294,265	0.422
	36	584,642	0.535
100 %	10	177,026	0.364
	18	328,419	0.461
	36	651,067	0.571

datos fueron factores clave para lograr este rendimiento. Estos aspectos permitieron una ejecución rápida sin comprometer la precisión ni la calidad de los resultados obtenidos. Esta eficiencia en el procesamiento es destacable, especialmente considerando la complejidad habitual de la reconstrucción tridimensional.

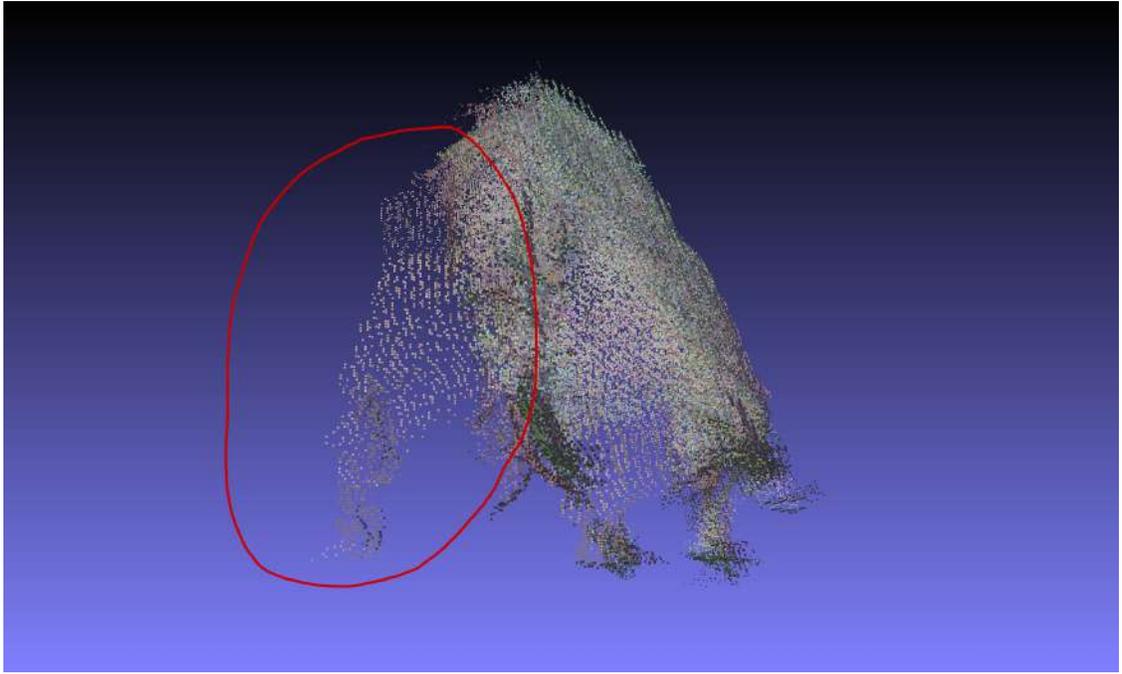


Figura 6.7: Error en ICP.

ICP y los transformers son herramientas poderosas en el campo del procesamiento de datos, pero se aplican a problemas muy diferentes. Compararlos directamente puede no ser del todo preciso, ya que operan en dominios distintos. Sin embargo, podemos analizar sus fortalezas y debilidades en sus respectivos contextos.

Ventajas del método ICP:

- **Precisión:** El ICP es un método preciso para la alineación de nubes de puntos, lo que permite obtener modelos 3D con un alto nivel de detalle.
- **Robustez:** El ICP es robusto frente al ruido y a las variaciones en la densidad de puntos, lo que lo hace adecuado para trabajar con datos RGB-D reales.
- **Eficiencia:** El ICP es un método computacionalmente eficiente, lo que permite procesar grandes cantidades de datos de manera rápida.
- **Versatilidad:** El ICP se puede aplicar a una amplia gama de objetos y entornos.
- **Facilidad de implementación:** El ICP es un método relativamente fácil de implementar,

lo que lo hace accesible para una amplia audiencia. El ICP es un método relativamente fácil de implementar, lo que lo hace accesible para una amplia audiencia.

Desventajas del método ICP:

- **Sensibilidad a valores atípicos:** El ICP puede ser sensible a valores atípicos en las nubes de puntos, lo que puede afectar la precisión de la alineación.
- **Dependencia de una estimación inicial precisa:** El ICP requiere una estimación inicial precisa de la alineación entre las nubes de puntos para funcionar correctamente.
- **Dificultades con superficies translúcidas:** El ICP puede tener dificultades para alinear superficies translúcidas, lo que limita su aplicabilidad en algunos casos.
- **Consumo de memoria:** El ICP puede consumir una gran cantidad de memoria, especialmente al trabajar con conjuntos de datos grandes.

Ventajas del método Geotransformer:

- **Captura de Relaciones de Largo Alcance:** Los Transformers son capaces de capturar dependencias de largo alcance dentro de los datos, lo que significa que pueden aprender relaciones complejas entre puntos distantes en la nube de puntos. Esto es particularmente útil cuando las nubes de puntos contienen grandes variaciones espaciales, algo que los métodos basados en ICP tienen dificultades para manejar.
- **Robustez ante Datos Ruidosos:** Los Transformers, especialmente en modelos avanzados como GeoTransformer, pueden ser más robustos frente a datos ruidosos y desordenados, gracias a sus mecanismos de atención que asignan pesos apropiados a diferentes partes de la nube de puntos. Esto los hace más resistentes a errores y ruidos en la adquisición de datos que los métodos tradicionales como ICP.
- **Ajuste Automático de la Geometría:** Los Transformers pueden aprender y ajustar automáticamente la geometría del objeto sin necesidad de intervención manual o de definir un número fijo de puntos de referencia. Esto hace que el proceso de alineación

y registro de nubes de puntos sea más flexible y menos dependiente de la configuración inicial.

- **Generalización a Escenarios Complejos:** Mientras que el método ICP requiere una configuración específica y es más adecuado para situaciones donde las nubes de puntos ya están cerca de una alineación inicial, los Transformers pueden generalizar mejor a situaciones complejas, como variaciones en la escala, rotaciones o deformaciones no lineales de los objetos.
- **Mejor manejo de Ambigüedades:** Los Transformers son buenos para manejar situaciones ambiguas o de baja resolución, ya que pueden combinar la información de varias vistas y ajustar la reconstrucción de acuerdo con las características aprendidas de los datos.

Desventajas del método Geotransformer:

- **Mayor Complejidad Computacional:** Los Transformers suelen ser más complejos en términos de arquitectura y requieren más recursos computacionales para su entrenamiento y ejecución. La necesidad de procesamiento paralelo y la atención a largo alcance pueden aumentar el tiempo de cálculo, especialmente en escenarios con grandes volúmenes de datos.
- **Requiere Gran Cantidad de Datos para Entrenamiento:** Los Transformers dependen en gran medida de grandes cantidades de datos etiquetados para su entrenamiento. Esto puede ser un desafío si no se dispone de suficientes datos de calidad para enseñar al modelo a realizar las tareas de reconstrucción de manera precisa.
- **Dependencia del Preprocesamiento:** Aunque los Transformers son robustos frente a datos ruidosos, aún dependen en gran medida del preprocesamiento de los datos. Si el preprocesamiento no se realiza adecuadamente (por ejemplo, si los datos no están correctamente alineados o filtrados), el rendimiento del modelo puede verse afectado.
- **Falta de Interpretabilidad:** A diferencia de los métodos basados en ICP, que son relativamente fáciles de entender y de ajustar manualmente, los modelos de Transformers,

al ser modelos de "caja negra", a menudo carecen de una interpretación clara sobre cómo están tomando las decisiones de alineación. Esto puede dificultar la identificación de errores en el proceso.

- **Requiere Afinación de Parámetros:** Aunque los Transformers pueden ser más flexibles, también requieren una cuidadosa afinación de los parámetros, como la tasa de aprendizaje, el número de capas de atención, entre otros. Si estos no se ajustan adecuadamente, el rendimiento del modelo puede no ser óptimo.

El ICP es ideal para escenarios en los que las nubes de puntos están inicialmente bien alineadas, permitiendo un ajuste preciso a nivel local. Es rápido y fácil de implementar, pero presenta limitaciones cuando se trata de datos ruidosos o grandes distorsiones geométricas.

Al superar estas limitaciones, el ICP puede convertirse en una herramienta aún más potente para la reconstrucción 3D a partir de datos RGB-D, con un amplio potencial para su aplicación en diversos campos.

En resumen, el ICP es un método valioso para la reconstrucción 3D utilizando datos RGB-D, que combina precisión, robustez, eficiencia y versatilidad. Sin embargo, es importante tener en cuenta sus limitaciones, como la sensibilidad a los valores atípicos y la necesidad de una estimación inicial precisa, para aprovecharlo de manera efectiva.

Por otro lado, los Transformers son especialmente adecuados para manejar nubes de puntos complejas y desordenadas. A diferencia del ICP, que se basa en un enfoque local para el ajuste de las nubes de puntos, los Transformers son modelos de aprendizaje profundo que pueden capturar patrones a largo alcance, lo que les permite ser más robustos frente a datos ruidosos y distorsiones geométricas complejas.

Los Transformers son capaces de aprender relaciones no solo entre puntos cercanos, sino también entre puntos distantes, lo cual es crucial en nubes de puntos dispersas o cuando se trata de objetos con geometrías complejas. Este enfoque permite que los Transformers mantengan una alta capacidad de generalización, incluso cuando los datos varían considerablemente. Además, a diferencia del ICP, que requiere una alineación inicial precisa, los Transformers pueden aprender patrones espaciales y temporales de manera más flexible, adaptándose a una mayor variedad de escenarios.

Sin embargo, esta flexibilidad viene con un costo: los Transformers requieren una mayor cantidad de datos para entrenar el modelo y, debido a su complejidad, demandan más recursos computacionales y tiempo de entrenamiento. A pesar de estos desafíos, su capacidad para manejar datos más complejos y aprender relaciones a gran escala les otorga una ventaja significativa en aplicaciones donde los datos son menos estructurados o contienen mucho ruido.

En general, los Transformers ofrecen una mayor capacidad de generalización y robustez frente a datos desordenados y complejos, mientras que el ICP sigue siendo más eficiente en escenarios más simples y con datos bien estructurados. Dependiendo de las características del problema y los datos, ambos métodos pueden complementarse o utilizarse por separado para obtener resultados óptimos en la reconstrucción 3D.

6.1. Proyección futura

El objetivo de los algoritmos desarrollados en este proyecto, junto con las adaptaciones pertinentes en el proceso de adquisición de datos, es sentar las bases para una reconstrucción tridimensional completa y precisa de los objetos, abarcando su geometría en su totalidad. Se espera que estos avances permitan generar una reconstrucción realista en tiempo real, optimizando tanto la calidad visual como la eficiencia computacional del sistema.

Algunos de los posibles usos y aplicaciones del proyecto son:

- Implementación del algoritmo para reconstrucción en tiempo real.
- Mejora de diseños protésicos generados.
- Optimización de objetos 3D para reducir materiales y aumentar la resistencia mecánica.
- Generación de nuevas estructuras con mayor resistencia mecánica.
- Reducción de los tiempos de generación de prototipos.
- Creación de meta-materiales para uso médico.

6.1. PROYECCIÓN FUTURA CAPÍTULO 6. RESULTADOS Y DISCUSIÓN

- Fomento del desarrollo de nuevas tecnologías en el campo del diseño de objetos 3D.
- Avance en la eficiencia y precisión de los procesos de diseño y fabricación.

Capítulo 7

Conclusiones

Las CNN han surgido como una alternativa prometedora al método ICP para el registro de nubes de puntos, ofreciendo ventajas significativas en cuanto a la precisión, robustez y eficiencia computacional. Sin embargo, la adopción generalizada de esta tecnología aún enfrenta algunos desafíos:

Desafíos:

- Interpretabilidad de los modelos: La naturaleza de “caja negra” de las CNN dificulta la comprensión de sus decisiones, lo que limita su uso en aplicaciones críticas.
- Disponibilidad de datos: Se requieren grandes conjuntos de datos de nubes de puntos alineadas con alta precisión para entrenar las CNN’s de manera efectiva.
- Generalización: Es crucial garantizar que las CNN entrenadas en un conjunto de datos específico puedan generalizarse a nuevos escenarios y tipos de datos.

A pesar de estos desafíos, las CNN’s presentan un enorme potencial para el registro de nubes de puntos, especialmente en el contexto de sensores RGB-D, donde pueden aprovechar la fusión de información multimodal y la robustez a las variaciones de iluminación.

Futuras investigaciones:

- Se deben desarrollar métodos para mejorar la interpretabilidad de los modelos de CNN, permitiendo una mayor confianza en sus resultados.
- Es necesario crear y recopilar conjuntos de datos de nubes de puntos alineadas de alta calidad para facilitar el entrenamiento de CNN's más robustas.
- Se deben explorar estrategias para mejorar la generalización de las CNN's a nuevos escenarios y tipos de datos, ampliando su aplicabilidad práctica.

En conclusión, las CNN's representan un avance significativo en el registro de nubes de puntos, pero aún requieren un mayor desarrollo para superar las limitaciones actuales y alcanzar su pleno potencial. La investigación continua en estas áreas es crucial para desbloquear todo el potencial de esta tecnología y abrir nuevas oportunidades en diversos campos como la realidad aumentada, la robótica y la conservación del patrimonio cultural.

Bibliografía

- [1] J. D. Fowler, *A History of Computer Graphics*. University Press of Florida, 2014.
- [2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, ser. Cambridge books online. Cambridge University Press, 2003.
- [3] S. Meerits, D. Thomas, V. Nozick, and H. Saito, “FusionMLS: Highly dynamic 3D reconstruction with consumer-grade RGB-D cameras,” *Computational Visual Media*, vol. 4, no. 4, pp. 287–303, 2018.
- [4] J. Liu, Y. Liu, and Z. Meng, “Point cloud registration leveraging structural regularity in manhattan world,” *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7888–7895, 2022.
- [5] Y. Wang and J. M. Solomon, “Deep closest point: Learning representations for point cloud registration,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 3523–3532.
- [6] P. J. Besl and N. D. McKay, “A method for registration of 3-d shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.
- [7] P. Henry, M. Krainin, E. Herbst, and X. Ren, “RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments,” *International Journal of Robotics Research*, vol. 31, no. 5, pp. 647–663, 2012.
- [8] Y. Chen and G. Medioni, “Object modeling by registration of multiple range images,” *Image and Vision Computing*, vol. 10, no. 3, pp. 145–155, 1992.

- [9] K. Doi, “Computer-Aided Diagnosis in medical imaging: historical review, current status and future potential,” *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, vol. 31, pp. 198–211, 2007.
- [10] M. D. Bermejillo Barrera, F. Franco-Martínez, and A. Díaz Lantada, “Artificial Intelligence Aided Design of Tissue Engineering Scaffolds Employing Virtual Tomography and 3D Convolutional Neural Networks,” *Materials*, vol. 14, no. 19, p. 5278, 2021.
- [11] C. Zhao, L. Sun, and R. Stolkin, “Simultaneous Material Segmentation and 3D Reconstruction in Industrial Scenarios,” *Frontiers in Robotics and AI*, vol. 7, 2020.
- [12] G. Zhang and Y. Chen, *Towards Optimal Point Cloud Processing for 3D Reconstruction*, 1st ed., ser. SpringerBriefs in Electrical and Computer Engineering. Springer Cham, 2022. [Online]. Available: <https://doi.org/10.1007/978-3-030-96110-7>
- [13] J. G. Bernal and otros, “Reducción de ruido por medio de redes neuronales,” 2008. [Online]. Available: <https://www.iiis.org/cds2008/cd2008csc/CISCI2008/PapersPdf/C836DN.pdf>
- [14] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Generative image inpainting with contextual attention,” 2018.
- [15] J. Smith and A. Doe, “Advanced 3d reconstruction techniques using rgb-d sensors,” *Journal of Computer Vision*, vol. 56, no. 4, pp. 123–142, 2022.
- [16] X. Wen, T. Li, Z. Han, and Y.-S. Liu, “Point cloud completion by skip-attention network with hierarchical folding,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1939–1948.
- [17] D. Maturana and S. Scherer, “VoxNet: A 3D Convolutional Neural Network for real-time object recognition,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2015, pp. 922–928.
- [18] B.-H. Li, B.-C. Hou, W.-T. Yu, and otros, “Applications of artificial intelligence in intelligent manufacturing: a review,” *Frontiers Inf Technol Electronic Eng*, vol. 18, p. 86–96, 2017.

- [19] L. Barbieri and M. Muzzupappa, “Performance-Driven Engineering Design Approaches Based on Generative Design and Topology Optimization Tools: A Comparative Study,” *Applied Sciences*, vol. 12, no. 4, p. 2106, 2022.
- [20] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation,” 2016.
- [21] D. Fofi, F. Remondino, and E. Nocerino, “State of the art and future challenges in RGB-D sensing for computer vision applications,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 104, pp. 92–116, 2015.
- [22] Microsoft, “Project Natal 101,” 2009, archivado desde el original el 21 de enero de 2012.
- [23] J. Han, L. Shao, D. Xu, and J. Shotton, “Enhanced computer vision with Microsoft Kinect sensor: A review,” *IEEE Transactions on Cybernetics*, vol. 43, no. 5, p. 1318–1334, 2013.
- [24] B. Aravind, S. Ajay, and Priyadarshini, “Virtualizing electrical appliances and gadgets input with human gestures,” *Procedia Computer Science*, vol. 50, pp. 503–510, 2015.
- [25] O. M. Salih, M. H. Rasheed, M. M. Siddeq, and M. A. Rodrigues, “Image compression for quality 3D reconstruction,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, 2022.
- [26] J. Li, W. Gao, Y. Wu, Y. Liu, and Y. Shen, “High-quality indoor scene 3D reconstruction with RGB-D cameras: A brief review,” 2022.
- [27] N. Wang and X. Gong, “Adaptive fusion for rgb-d salient object detection,” *IEEE Access*, vol. 7, 2019.
- [28] S. Shah and D. J. Kriegman, “Depth from focus using RGB-D cameras,” in *Computer Vision and Pattern Recognition Workshops*, 2012, pp. 1–7.
- [29] M. B. Shaikh and D. Chai, “Rgb-d data-based action recognition: A review,” 2021.

- [30] D. Azinović, R. Martin-Brualla, D. B. Goldman, M. Nießner, and J. Thies, “Neural RGB-D surface reconstruction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6290–6301.
- [31] H. Kim and T. Nguyen, “Comparative analysis of rgb-d sensor technologies in 3d reconstruction,” *Sensors and Actuators A: Physical*, vol. 299, p. 112034, 2022.
- [32] Intel®, “Depth camera D415,” Intel® RealSense™ Depth and Tracking Cameras., 2023. [Online]. Available: <https://www.intelrealsense.com/depth-camera-d415/>
- [33] K. Lee and R. Patel, “Iterative closest point algorithm: Applications in robotic mapping and 3d modeling,” *International Journal of Robotics Research*, vol. 45, no. 2, pp. 75–95, 2021.
- [34] Y. He, B. Liang, J. Yang, S. Li, and J. He, “An iterative closest points algorithm for registration of 3d laser scanner point clouds with geometric features,” *Sensors*, vol. 17, no. 8, p. 1862, 2017.
- [35] J. Zhang, Y. Yao, and B. Deng, “Fast and robust iterative closest point,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1–1, 2021. [Online]. Available: <http://dx.doi.org/10.1109/TPAMI.2021.3054619>
- [36] G. W. Flake, *The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation*. MIT Press, 1998.
- [37] B.-K. Sy, “An AI-based CAD/CAM approach to assess design methodology of a user-specific nonvocal communication device,” in *Proceedings of the Fifteenth Annual Northeast Bioengineering Conference*, 1989, pp. 185–186.
- [38] R. M. Golden and R. Golden, *Mathematical methods for neural network analysis and design*. MIT Press, 1996.
- [39] B. hu Li, B. cun Hou, W. tao Yu, X. bing Lu, and C. wei Yang, “Applications of artificial intelligence in intelligent manufacturing: a review,” *Frontiers of Information Technology and Electronic Engineering*, vol. 18, no. 1, pp. 86–96, 2017.

- [40] D. S. Sabbella, A. Singh, and U. G. Maheswari, “Artificial intelligence in 3D CAD modelling,” in *International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 2020, pp. 1–5.
- [41] M. A. Aceves Fernández, *Inteligencia Artificial Para Programadores Con Prisa*. Universo de Letras, 2021.
- [42] O. Theobald, *Machine Learning for Absolute Beginners: A Plain English Introduction*, 2nd ed. Scatterplot Press, 2017, vol. 3. [Online]. Available: <https://books.google.com.mx/books?id=PGNzswEACAAJ>
- [43] L. Luchesi, R. O. Foschi, M. Rubinstein, and Möller, “Redes neuronales aplicadas a dinámica de estructuras,” *Mecánica Computacional*, vol. XXIV, pp. 797–812, 2005.
- [44] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2023.
- [45] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” 2016.
- [46] Google AI, “Google translate,” <https://translate.google.com/>, [Accessed: 22-ago-2024].
- [47] —, “Google ai text summarization,” <https://ai.google/research/pubs/>, [Accessed: 22-ago-2024].
- [48] —, “Lamda: Language model for dialogue applications,” <https://ai.googleblog.com/2021/05/lamda-towards-safe-grounded-and-high.html>, [Accessed: 22-ago-2024].
- [49] T. Brown, B. Mann, N. Ryder *et al.*, “Language models are few-shot learners,” *Advances in Neural Information Processing Systems*, vol. 33, 2020. [Online]. Available: <https://arxiv.org/abs/2005.14165>
- [50] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-end object detection with transformers,” *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. [Online]. Available: <https://arxiv.org/abs/2005.12872>

- [51] R. Flóres Lopez and J. M. Fernández Fernández, *Las redes Neuronales Artificiales. Metodología y Análisis de datos en Ciencias Sociales*. Netbiblo, 2008.
- [52] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, “Meshlab: an open-source mesh processing tool,” in *Sixth Eurographics Italian Chapter Conference*, 2008, pp. 129–136.
- [53] T. Jost and H. Hügli, “Fast ICP Algorithms for Shape Registration,” in *Springer Book Archive*, 2002, pp. 91–99.
- [54] J. Zhang, Y. Yao, and B. Deng, “Fast and robust iterative closest point,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3450–3466, 2022.
- [55] J. M. Aviña Muñoz, J. C. Pedraza Ortega, J. M. Ramos Arreguín, and L. R. Román Rivera, “Evaluación de métodos iterativos de registro de nubes de puntos y su uso en reconstrucción 3D de objetos.” *La Mecatrónica en México*, vol. 13, no. 2, pp. 40–58, May 2024. [Online]. Available: www.mecamex.net/revistas/LMEM
- [56] Z. Qin, H. Yu, C. Wang, Y. Guo, Y. Peng, and K. Xu, “Geometric transformer for fast and robust point cloud registration,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 11 143–11 152.
- [57] M. A. Uy, Q.-H. Pham, B.-S. Hua, D. T. Nguyen, and S.-K. Yeung, “Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data,” in *International Conference on Computer Vision (ICCV)*, 2019.
- [58] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, “3d shapenets: A deep representation for volumetric shapes,” in *Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, oral Presentation. [Online]. Available: <https://shapenets.cs.princeton.edu/>
- [59] S. Huang, Z. Gojcic, M. Usvyatsov, A. Wieser, and K. Schindler, “Predator: Registration of 3d point clouds with low overlap,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 4267–4276.

- [60] Z. J. Yew and G. H. Lee, “RPM-Net: Robust Point Matching Using Learned Features,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 824–11 833.
- [61] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “Pointnet: Deep learning on point sets for 3d classification and segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 652–660.
- [62] Image to STL, “Convert ply to off,” <https://imagetostl.com/convert/file/ply/to/off#convert>, accessed: 2024-10-14.
- [63] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas, “Kpconv: Flexible and deformable convolution for point clouds,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 6411–6420.
- [64] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2117–2125.
- [65] H. Yu, F. Li, M. Saleh, B. Busam, and S. Ilic, “Cofinet: Reliable coarse-to-fine correspondences for robust point cloud registration,” *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, 2021.
- [66] J. Li, B. M. Chen, and G. H. Lee, “So-net: Self-organizing network for point cloud analysis,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 9397–9406.
- [67] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, “SuperGlue: Learning Feature Matching with Graph Neural Networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 4938–4947.
- [68] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, “Pytorch: An imperative style, high-performance deep

learning library,” in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019, pp. 8026–8037.

- [69] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2015.

Capítulo 8

Anexos

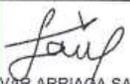
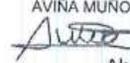
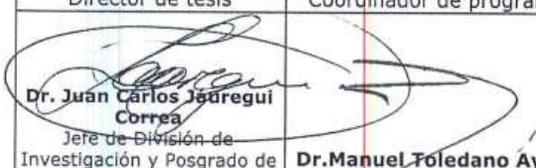
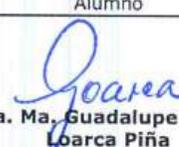
A continuación se muestran los anexos correspondientes para la obtención del grado de Maestro en Ciencias en Inteligencia Artificial.



DIPFI
POSGRADO
INGENIERÍA

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE INGENIERÍA
DIVISIÓN DE INVESTIGACIÓN Y POSGRADO

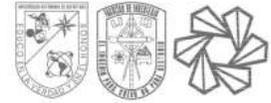
**REGISTRO DEL PROTOCOLO DE INVESTIGACIÓN
DEL ESTUDIANTE DE POSGRADO**

Los 2 Espacios oscuros exclusivos para la Dirección	No. Registro de Proyecto*:	14273
	Fecha de Registro*:	24/10/2023
	Fecha de inicio de proyecto:	01/08/2022
	Fecha de término de proyecto:	31/07/2024
1. DATOS DEL SOLICITANTE		
No. de expediente:	317937	
Apellido Paterno	Apellido Materno	Nombre(s)
AVIÑA	MUÑOZ	JUAN MANUEL
Dirección:		
Calle y número	Colonia	CP
Senda Eterna #626	Milenio III	76080
Estado	Teléfono (incluír lada)	Correo Electrónico
22	4422499033	JuanAvina.m@uaq.mx
2. DATOS DEL PROYECTO		
Facultad:	INGENIERÍA	
Programa:	MAESTRIA EN CIENCIAS EN INTELIGENCIA ARTIFICIAL	
Tema específico del proyecto:	Reconstrucción de objetos 3D usando cámaras RGB-D y Redes Neuronales	
 PEDRAZA ORTEGA JESUS CARLOS Director de tesis	 TOVAR ARRIAGA SAUL Coordinador de programa	 AVIÑA MUÑOZ JUAN MANUEL Alumno
 Dr. Juan Carlos Jauregui Correa Jefe de División de Investigación y Posgrado de la Fac. de Ing.	 Dr. Manuel Toledano Ayala Director de Fac. Ing.	 Dra. Ma. Guadalupe Flavia Loarca Piña Directora de Investigación y Posgrado UAQ



Anexo 3 Manual de procedimientos Administrativos de Posgrado de la Facultad de Ingeniería

Figura 8.1: Registro de protocolo.



C. U., 9 de septiembre de 2023

Juan Manuel Aviña Muñoz
Estudiante de Maestría en Ciencias, Inteligencia Artificial
Expediente 317937

Presente

El Comité de Ética Aplicada a la Investigación de la Facultad de Ingeniería ha revisado el protocolo del trabajo de tesis:

CEAIFI-129-2023-TP

**Reconstrucción de objetos 3D usando cámaras RGB-D
y Redes Neuronales**

Con apego a los lineamientos éticos de beneficencia, no maleficencia, justicia y autonomía, este comité ha dado el siguiente dictamen:

Exento de dictamen ético

El presente dictamen tiene vigencia de un año a partir de su fecha de emisión.

Sirva esta carta para los fines académicos que al interesado convengan.

Atentamente

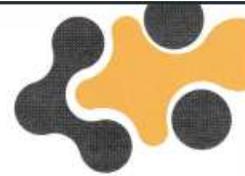
“El ingenio para crear, no para destruir”

Dra. Aurora Femat Díaz
Presidente del CEAIFI
afemat@uaq.mx

Figura 8.2: Comité de Ética.



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE LENGUAS Y LETRAS



A QUIEN CORRESPONDA:

La que suscribe, Directora de la Facultad de Lenguas y Letras, hace **C O N S T A R** que

AVIÑA MUÑOZ JUAN MANUEL

Presentó y acreditó el **Examen de Comprensión de Textos en Inglés** efectuado el día veinticinco de septiembre de dos mil veinticuatro.

Se extiende la presente a petición de la parte interesada, para los fines escolares y legales que le convengan, en el Campus Aeropuerto de la Universidad Autónoma de Querétaro, el día quince de octubre de dos mil veinticuatro.

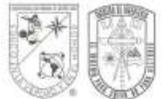


Atentamente,
 "Enlazar Culturas por la Palabra"


DRA. MA. DE LOURDES RICO CRUZ

MLRC/mgoa*CL*FLL-C.-133

Figura 8.3: Constancia Nivel de Inglés.



SECRETARÍA ADMINISTRATIVA
FACULTAD DE INGENIERÍA

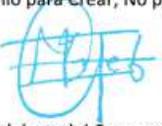
QUIEN CORRESPONDA:

Por este conducto, hago constar que el (la) C. **Juan Manuel Aviña Muñoz** con número de expediente **317937** de **MAESTRÍA EN CIENCIAS EN INTELIGENCIA ARTIFICIAL** de la Facultad de Ingeniería de la UAQ, **No tiene ADEUDO** en:

- SET DE TV VIRTUAL
- LABORATORIO MULTIMEDIA INGENIA
- LABORATORIO DE AUTOMATIZACIÓN
- LABORATORIO DE MECATRÓNICA
- LABORATORIO DE FÍSICA
- LABORATORIO DE QUÍMICA
- LABORATORIO DE GEOTECNIA AMBIENTAL
- LABORATORIO DE HIDRÁULICA
- LABORATORIO DE GEOTECNIA, MATERIALES Y GEOMÁTICA
- LABORATORIO DE PROTOTIPOS (PROTOLAB)
- LABORATORIO DE CERÁMICA
- LABORATORIO DE INNOVACIÓN SOCIAL
- LABORATORIO DE DISEÑO DE EQUIPO MEDICO

A petición del (a) interesada (o) se extiende la presente en el Centro Universitario de la ciudad de **Santiago de Querétaro, Qro., 15/10/2024**, para los trámites académicos que se requieran.

Atentamente
"El Ingenio para Crear, No para Destruir"



C.P. Guadalupe del Carmen Molinero González
Secretaria Administrativa FI de la UAQ

Folio 1527
VIGENCIA 30 DÍAS NATURALES A PARTIR DE LA EMISIÓN.

Ccp. Area correspondiente
ARCHIVO

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
CERRO DE LAS CAMPANAS S/N, COL. LAS CAMPANAS
C.P. 76010 QUERÉTARO, QRO.
TEL: (442) 192-12-00 EXT. 6017

Figura 8.4: Constancia no adeudo Laboratorios.



Centro Universitario, a 07 de noviembre de 2024.

M. en C. Darío Hurtado Maldonado
 Director de Servicios Académicos
 Presente.

Por este medio informo a usted que el candidato **Juan Manuel Aviña Muñoz**, con expediente: 317937 egresado del programa: **Maestría en Ciencias en Inteligencia Artificial**, con la opción de obtención de grado por elaboración y defensa de tesis, cumplió con los siguientes requisitos:

- Guías y formatos establecidos por la Dirección de Investigación y Posgrado, de acuerdo con la opción autorizada.
- Liberación de Adeudos de la facultad.
- Requisitos adicionales de la facultad.

Por lo anterior, se autoriza la integración de su expediente ante la Dirección de Servicios Académicos para que sea turnado al H. Consejo Universitario.

Sin otro particular, reciba un cordial saludo.

Atentamente



Dr. Juan Carlos Lauregui Correa,
 Jefe de la DIPFI.

"El Ingenio para crear, no para destruir"



Figura 8.5: Carta de Requisitos.



Evaluación del método ICP para registro de nubes de puntos y su uso en reconstrucción 3D de objetos

Aviña Muñoz Juan Manuel (✉), Pedraza Ortega Jesús Carlos, Ramos Arreguín Juan Manuel, Román Rivera Luis Rogelio

Facultad de Ingeniería, Universidad Autónoma de Querétaro, México
✉juan.avina.m@gmail.com, caryoko@yahoo.com, jsistdig@yahoo.com.mx, r3roger@gmail.com

Resumen

El presente estudio tiene como objetivo desarrollar un método preciso para la reconstrucción 3D de objetos a partir de datos adquiridos por sensores RGB-D. La metodología propuesta integra el método Iterative Closest Point (ICP) para la fusión de nubes de puntos, permitiendo obtener modelos 3D detallados y precisos. Se describe el proceso de captura de datos, procesamiento y generación del modelo 3D, destacando los desafíos y soluciones encontrados en la alineación de las nubes de puntos. Los resultados obtenidos demuestran la efectividad del método ICP para generar modelos 3D precisos, con un impacto potencial en campos como la ingeniería, el diseño y la medicina. Este trabajo sirve como referencia para analizar la eficiencia de este método a diferentes densidades en las nubes de puntos.

Palabras clave: Reconstrucción 3D, Cámaras RGB-D, ICP, métodos iterativos, Registro de Nubes de Puntos, Modelos 3D.

Abstract

The present study aims to develop an accurate method for 3D reconstruction of objects from data acquired by RGB-D sensors. The proposed methodology integrates the Iterative Closest Point (ICP) method for point cloud fusion, allowing to obtain detailed and accurate 3D models. The process of data capture, processing and generation of the 3D model is described, highlighting the challenges and solutions found in the alignment of the point clouds. The results obtained demonstrate the effectiveness of the ICP method to generate accurate 3D models, with a potential impact in fields such as engineering, design and medicine. This work serves as a reference to analyze the efficiency of this method at different point cloud densities.

Keywords: 3D Reconstruction, RGB-D Cameras, ICP, iterative methods, Point Cloud Registration, 3D Models.

1. Introducción

La reconstrucción de objetos 3D, esa increíble habilidad de dar vida a objetos del pasado o de plasmar ideas en modelos digitales, no es un invento reciente. Su historia se remonta a tiempos remotos, donde la creatividad y el ingenio humano se unían a las herramientas disponibles para dar forma a la visión.

Las primeras técnicas de reconstrucción 3D se basaban en métodos manuales y artesanales. En

Figura 8.6: Portada del Artículo en LMeM.

La Mecatrónica en México



DICTÁMEN DE EVALUACIÓN

3 de Mayo de 2024

OFICIO LMEM/20240503-1

Por este medio, el comité editorial de la revista LMEM se les informa a los autores que el trabajo titulado

Evaluación del método ICP para registro de nubes de puntos y su uso en reconstrucción 3D de objetos

Con número de registro **LMEM20240401-1**, cuyos autores son

Aviña Muñoz Juan Manuel, Pedraza Ortega Jesús Carlos, Ramos Arreguín Juan Manuel, Román Rivera Luis Rogelio

Ha tenido como dictamen **APROBADO**.

Por lo que se invita a los autores a continuar con el proceso de publicación, aprobando el borrador, enviando la carta de autorización de publicación, y realizando el pago correspondiente.

ATENTAMENTE,
Comité Evaluador de los Artículos de la revista
"La Mecatrónica en México"



ASOCIACIÓN MEXICANA
DE MECATRÓNICA A.C.

Figura 8.7: Dictamen del Artículo en LMeM.

La Mecatrónica en México



DECLARACIÓN DE DERECHOS DE PUBLICACIÓN

Por este medio, los autores del artículo titulado

Evaluación del método ICP para registro de nubes de puntos y su uso en reconstrucción 3D de objetos

Conceden los derechos de publicación del artículo mencionado en la Revista La Mecatrónica en México.

Los autores declaran no tener conflictos de interés, y que su artículo no ha sido publicado en otro medio, ni está en proceso de revisión para ser publicado, aparte de esta revista. Así mismo deslindamos a la revista de cualquier conflicto que pudiera existir por similitudes con otros trabajos.

Juan Manuel Aviña Muñoz		12/05/2024
Autor 1	Firma	Fecha
Jesús Carlos Pedraza Ortega		12/05/2024
Autor 2	Firma	Fecha
Juan Manuel Ramos Arreguín		12/05/2024
Autor 3	Firma	Fecha
Luis Rogelio Román Rivera		12/05/2024
Autor 4	Firma	Fecha
Autor 5	Firma	Fecha



ASOCIACIÓN MEXICANA DE MECATRÓNICA A.C.

Figura 8.8: Derechos de Publicación del Artículo en LMeM.