

Sistema de navegación en interiores asistido por visión computacional orientado a personas con
discapacidad visual

Manuel Stiven Leon Males
Luis Felipe Paz Caicedo

Universidad CESMAG
Facultad de Ingeniería
Programa de Ingeniería Electrónica
San Juan de Pasto
2026

Sistema de navegación en interiores asistido por visión computacional orientado a personas con
discapacidad visual

Manuel Stiven Leon Males

Luis Felipe Paz Caicedo

Proyecto de trabajo de grado presentado al Comité Curricular del Programa de Ingeniería
Electrónica en la modalidad Investigación

Asesor

Mg. José Camilo Eraso Guerrero

Universidad CESMAG
Facultad de Ingeniería
Programa de Ingeniería Electrónica
San Juan de Pasto

2026

Nota de aceptación

Aprobado por el Comité de Grado en cumplimiento de los requisitos exigidos por la Universidad CESMAG para optar por el título de Ingeniero Electrónico.

Ing. José Camilo Eraso Guerrero

Asesor

Jurado

Jurado

San Juan de Pasto, octubre 2025

Nota de exclusión

“El pensamiento que se expresa en esta obra es exclusiva responsabilidad del autor y no compromete la ideología de la Universidad CESMAG”.

Dedicatoria

A mi madre, por ser mi mayor apoyo, mi inspiración y mi refugio en los momentos difíciles. Gracias por cada sacrificio silencioso, por cada palabra de aliento y por enseñarme, con su ejemplo, que la constancia y la humildad abren camino incluso en las circunstancias más difíciles. Todo lo que hoy alcanzo lleva una parte de usted.

Manuel Stiven Leon Males

A mis padres, cuyo apoyo incondicional, comprensión y sacrificio hicieron posible alcanzar este logro. Gracias por acompañarme en cada etapa de mi formación y por impulsarme a continuar con determinación. Todo lo que he logrado en este camino es reflejo de los valores que me inculcaron y del amor constante que siempre me han brindado.

Luis Felipe Paz Caicedo

Agradecimientos

Expresamos nuestro sincero agradecimiento a la Universidad, por brindarnos la formación académica, las herramientas necesarias y el acompañamiento constante a lo largo de nuestro proceso profesional. Su compromiso con la calidad educativa representó un pilar fundamental en nuestro desarrollo, impulsándonos a asumir cada reto con dedicación y responsabilidad.

Extendemos también nuestro reconocimiento a cada una de las personas que integran la comunidad universitaria, cuya disposición, apoyo y compromiso aportaron significativamente a nuestra formación tanto en el ámbito académico como en el personal. Cada contribución, desde sus diferentes roles, dejó una huella importante en este camino.

De manera especial, agradecemos a los docentes del programa de Electrónica, por compartir sus conocimientos, experiencias y enseñanzas con vocación, exigencia y entrega. Su guía fue esencial para fortalecer nuestras capacidades, ampliar nuestra visión profesional y motivarnos a continuar aprendiendo más allá de las aulas.

De manera muy especial, queremos expresar nuestra gratitud al Ingeniero José Camilo Eraso Guerrero, por su acompañamiento, orientación y asesoría durante el desarrollo de este trabajo de grado. Gracias a su asesoramiento, compromiso, experiencia y constante disposición, fue posible encaminar adecuadamente esta investigación, superar los desafíos presentados y alcanzar satisfactoriamente los objetivos propuestos.

Contenido

Lista de Tablas	12
Lista de Figuras	13
Introducción	15
1 El problema de investigación	16
1.1 Objeto o tema de investigación	16
1.2 Línea de investigación.....	16
1.3 Sub-línea de Investigación.	16
1.4 Planteamiento o Descripción del Problema	16
1.5 Formulación del Problema	18
1.6 Objetivos	18
1.6.1 Objetivo General	18
1.6.2 Objetivos Específicos.....	18
1.7 Justificación.....	19
1.8 Viabilidad	20
1.9 Delimitación	21
2 Tópicos del Marco Teórico.....	22
2.1 Antecedentes	22
2.1.1 Detección de objetos en tiempo real con YOLO	22
2.1.2 Detección híbrida de objetos para asistencia visual.....	23
2.1.3 Avances Recientes en Aplicaciones y Métodos de Aprendizaje Profundo para la Navegación Autónoma: Una Revisión Integral	24
2.1.4 Silla de ruedas inteligente para percepción y evitación de obstáculos.....	25
2.1.5 Caminador inteligente para asistencia de movilidad.....	25
2.1.6 Visión Estéreo en Asistencia para la Navegación de Personas Ciegas	26

2.1.7	Visión Artificial para Personas con Discapacidad Visual utilizando Detección con YOLO	27
2.1.8	Conversión de Imágenes a Voz para Personas con Discapacidad Visual	28
2.1.9	Detección de Objetos en Tiempo Real y Retroalimentación de Audio para Personas con Discapacidad Visual.....	28
2.1.10	Third Eye: Reconocimiento de Objetos y Generación de Habla para Personas con Discapacidad Visual.....	30
2.2	Enunciados de los supuestos teóricos.....	33
2.2.1	Visión computacional.....	33
2.2.2	Redes neuronales convolucionales (CNN).....	34
2.2.3	Algoritmos de Deep Learning para visión computacional.....	36
2.2.4	Estereovisión	41
2.2.5	Discapacidad visual.....	42
2.2.6	Sistema de navegación	42
2.2.7	Salud visual	42
2.2.8	Entornos interiores	42
2.2.9	Sistemas de procesamientos de datos.....	43
2.2.10	Baja visión.....	43
2.2.11	Retroalimentación auditiva	44
2.2.12	Modelos pre entrenados	44
2.2.13	Bounding Box	45
2.2.14	Sistema completo (end-to-end)	45
2.3	Definición de conceptos	45
2.3.1	Definición nominal de conceptos.....	45
2.3.2	Definición operativa de conceptos	46
2.4	Hipótesis.....	49

2.4.1	Hipótesis de investigación.....	49
2.4.2	Hipótesis nula.....	50
2.4.3	Hipótesis alternativa.....	50
3	Metodología.....	51
3.1	Enfoque	51
3.2	Paradigma.....	51
3.3	Método	52
3.4	Tipo de investigación	52
3.5	Diseño de investigación	52
3.6	Universo	53
3.7	Muestra.....	53
3.8	Técnicas de recolección de información	54
3.8.1	Validez de la técnica	54
3.8.2	Confiabilidad técnica	54
3.9	Instrumentos de recolección de información	55
3.10	Establecimiento de requerimientos para el sistema.....	55
3.11	Diseño del sistema.....	56
3.11.1	Plataforma de hardware utilizada.....	56
3.11.2	Diseño 3D del soporte para el módulo de cámaras estéreo.....	56
3.11.3	Conexión del módulo de doble lente con el computador.....	57
3.11.4	Configuración del módulo de visión.....	57
3.11.5	Diseño del dataset	58
3.11.6	Entrenamiento del modelo YOLOv8	61
3.11.7	Diseño de la retroalimentación auditiva.....	62
3.11.8	Diseño de las instrucciones.....	62

3.11.9	Configuración inicial del módulo del sistema.....	65
3.12	Implementación del sistema	65
3.12.1	Integración del hardware.....	65
3.12.2	Integración de software.....	66
3.12.3	Interfaz de monitoreo	67
3.12.4	Configuración del prototipo y modo de operación	67
3.13	Protocolo experimental y criterios de evaluación.....	68
3.13.1	Criterios de evaluación por módulo	68
4	Resultados.....	71
4.1	Prototipo del módulo de cámaras instalado en el sistema.....	71
4.2	Resultados de detección del modelo	71
4.3	Resultados del entrenamiento del modelo.....	73
4.4	Resultados experimentales del sistema	76
4.4.1	Escenario 1: Salón H501	76
4.4.2	Escenario 2: Laboratorio de Física.....	89
4.4.3	Escenario 3: Sexto piso – Edificio San Francisco.....	92
4.4.4	Evaluación comparativa para análisis de eficiencia.....	97
4.4.5	Eficiencia del sistema.....	99
4.4.6	Velocidad (FPS).....	101
4.4.7	Velocidad de emisión de instrucciones gTTS.....	101
4.4.8	Costo computacional.....	102
4.4.9	Eficacia del sistema.....	102
4.4.10	Efectividad global del sistema	104
5	Análisis y discusión de resultados	105
5.1	Consideraciones generales del análisis	105

5.2	Análisis de resultados de detección de instancias (criterio A).....	106
5.3	Análisis de resultados de clasificación de objetos (criterio B).....	107
5.4	Análisis de la estimación de distancia con estereovisión.....	108
5.4.1	Justificación del umbral ± 10 cm.....	109
5.4.2	Análisis del comportamiento de la estereovisión.....	109
5.5	Análisis del criterio D: Asignación de zona espacial.....	110
5.5.1	Consistencia algorítmica.....	110
5.5.2	Robustez en zonas limítrofes.....	110
5.5.3	Independencia de factores ambientales.....	111
5.6	Análisis del sistema completo (end-to-end).....	111
5.6.1	Desempeño comparativo por escenario.....	112
5.6.2	Patrones críticos identificados.....	112
5.6.3	Jerarquía crítica de módulos.....	113
5.6.4	Análisis de evaluación comparativa de navegación con y sin sistema.....	114
5.6.5	Análisis de la eficiencia del sistema de navegación asistida.....	114
5.7	Análisis de la velocidad de procesamiento (FPS).....	115
5.8	Análisis del sistema de retroalimentación auditiva.....	116
5.9	Análisis del costo computacional.....	116
5.10	Análisis de la efectividad global del sistema.....	116
6	Conclusiones.....	118
7	Recomendaciones.....	121
8	Trabajos futuros.....	123
	Referencias.....	125
9	Anexos.....	129

Lista de Tablas

Tabla 1 <i>Principal aporte de investigación</i>	31
Tabla 2 <i>Clases del dataset utilizadas para el entrenamiento del modelo YOLOV8n</i>	59
Tabla 3 <i>Métricas de desempeño por clase para el sistema de detección modular YOLOv8n</i>	74
Tabla 4 <i>Detección de objetos (Criterio A)</i>	78
Tabla 5 <i>Desempeño de clasificación de objetos (Criterio B)</i>	81
Tabla 6 <i>Precisión de la estimación de distancias (Estereovisión) sobre objetos detectados (Criterio C)</i>	84
Tabla 7 <i>Resultados de la validación de asignación de zonas espaciales (Criterio D)</i>	85
Tabla 8 <i>Resultados de la evaluación del sistema de navegación por visión computacional (YOLO + estereovisión)</i>	87
Tabla 9 <i>Métricas globales del sistema completo (end-to-end)</i>	89
Tabla 10 <i>Desempeño del sistema por criterios individuales – Escenario 2 (Laboratorio de Física)</i>	92
Tabla 11 <i>Resultados porcentuales del sistema – Escenario 2</i>	92
Tabla 12 <i>Desempeño del sistema por criterios individuales – Escenario 3 (Sexto piso del edificio San Francisco)</i>	96
Tabla 13 <i>Resultados porcentuales del sistema – Escenario 3</i>	96
Tabla 14 <i>Resultados de navegación sin asistencia</i>	98
Tabla 15 <i>Resultados de navegación con asistencia del sistema</i>	99
Tabla 16 <i>Comparación de navegación con y sin asistencia</i>	99
Tabla 17 <i>Eficacia global del sistema de navegación asistida por escenario</i>	103
Tabla 18 <i>Recall del módulo de detección por escenario experimental</i>	106
Tabla 19 <i>Accuracy del módulo de clasificación por escenario experimental</i>	108
Tabla 20 <i>Muestra los valores promedio de Precisión, Recall y F1-Score obtenidos en cada escenario durante las pruebas experimentales</i>	111

Lista de Figuras

Figura 1 <i>Esquema conceptual de un sistema de detección de objetos con retroalimentación auditiva para personas con discapacidad visual basado en YOLO.</i>	22
Figura 2 <i>Arquitectura general de un sistema de navegación asistida mediante detección de objetos basada en YOLOv8.</i>	24
Figura 3 <i>Esquema conceptual del uso de visión estéreo para la estimación de distancia y detección de obstáculos en sistemas de navegación asistida.</i>	26
Figura 4 <i>Comparación conceptual entre el uso de un conjunto de datos general y un conjunto de datos específico del entorno para la detección de objetos.</i>	29
Figura 5 <i>Convolución discreta en 2D — Dumoulin y Visin [15]</i>	35
Figura 6 <i>Average Pooling 5×5 — Dumoulin y Visin [15]</i>	36
Figura 7 <i>División en cuadrícula y predicción de bounding boxes — Redmon et al. [16]</i>	38
Figura 8 <i>Arquitectura completa YOLOv8 — Yaseen [17]</i>	40
Figura 9 <i>Modelo YOLOv8 en la detección, segmentación, clasificación y estimación de postura.</i> [19].....	41
Figura 10. <i>Diseño 3D del soporte</i>	57
Figura 11 <i>Patrón captado por ambos lentes (izquierdo y derecho) para calibración de estereovisión</i>	58
Figura 12 <i>Ejemplo de objeto siendo etiquetado con LabelImg</i>	60
Figura 13 <i>Audífonos inalámbricos de conducción aérea.</i>	62
Figura 14. <i>Diagrama de flujo de generación de mensajes de voz.</i>	64
Figura 15 <i>Diagrama de flujo de conexión entre hardware y software.</i>	66
Figura 16 <i>Configuración física del prototipo en su diseño portátil</i>	66
Figura 17 <i>Módulo de doble lente para estereovisión junto a la correa y el soporte</i>	71
Figura 18 <i>Resultados de detección obtenidos durante la validación</i>	72
Figura 19 <i>Curvas de rendimiento del modelo YOLOv8n durante la fase de validación</i>	75
Figura 20 <i>Vista panorámica del entorno experimental: Salón H501 (Escenario 1)</i>	77
Figura 21 <i>Esquema del escenario 1 (Salón H501) con la ruta de desplazamiento y disposición de objetos</i>	77
Figura 22 <i>Ejemplo de objeto detectado (TP)</i>	80
Figura 23 <i>Fallo de detección (FN): silla presente pero no detectada. Colisión en prueba 5</i>	81

Figura 24 <i>Validación exitosa de los criterios de clasificación, distancia y zona</i>	83
Figura 25 <i>Error de clasificación: bote de basura etiquetado incorrectamente como mueble</i>	83
Figura 26 <i>Evaluación del criterio C. La captura ilustra una imprecisión en la estimación de distancia</i>	85
Figura 27 <i>Validación geométrica de la asignación de zona en el Escenario 1 (ejemplo representativo)</i>	87
Figura 28 <i>Vista panorámica del entorno experimental: Laboratorio de Física (Escenario 2)</i> ...	90
Figura 29 <i>Esquema del escenario 2 (Laboratorio de Física) con la ruta de desplazamiento y disposición de objetos</i>	90
Figura 30 <i>Vista panorámica del entorno experimental: Sexto piso del edificio San Francisco (Escenario 3)</i>	94
Figura 31 <i>Esquema del escenario 2 (Laboratorio de Física) con la ruta de desplazamiento y disposición de objetos</i>	95
Figura 32 <i>Ruta controlada, diseñada para realizar las pruebas de comparación</i>	98
Figura 33 <i>Caso crítico de detección parcial: identificación correcta de un objeto secundario y omisión de un obstáculo frontal prioritario</i>	107

Introducción

La movilidad en interiores, como oficinas o instalaciones universitarias, supone grandes desafíos para personas con discapacidad visual. La disposición cambiante de muebles y el flujo constante de personas dificultan la orientación segura. Escaleras, entradas y pasillos estrechos aumentan los riesgos, y aunque herramientas convencionales de asistencia, como el bastón blanco, así como tecnologías de posicionamiento como el GPS, representan apoyos fundamentales para la movilidad, presentan limitaciones en entornos interiores complejos, especialmente al no ofrecer información contextual en tiempo real ni reconocimiento de objetos cercanos, lo que puede reducir la autonomía del usuario en escenarios dinámicos.

Para abordar estos retos, como complemento a las herramientas convencionales de asistencia, un sistema inteligente de navegación asistido por visión computacional ofrece una alternativa tecnológica orientada a enriquecer la percepción del entorno. Basado en algoritmos como YOLOv8 para el reconocimiento de objetos y estereovisión con OpenCV, este sistema identifica obstáculos y señales, e integra alertas de voz mediante tecnología text-to-speech que informan en tiempo real sobre obstáculos y rutas seguras, contribuyendo a mejorar la orientación, autonomía y seguridad del usuario en espacios interiores complejos.

Este sistema tiene potencial de aplicación en entornos interiores como edificios públicos, oficinas y áreas industriales cerradas, así como en el apoyo complementario a personas con discapacidad visual. Su implementación en plataformas portátiles permitiría proporcionar retroalimentación auditiva en tiempo real para alertar sobre obstáculos y fortalecer la orientación espacial del usuario.

La presente investigación tiene como objetivo evaluar la efectividad de un sistema de navegación en interiores, orientado a personas con discapacidad visual, que utiliza visión computacional para el reconocimiento de objetos y la estimación de proximidad en tiempo real. Basado en el modelo YOLOv8 y la librería OpenCV, el sistema se implementará en la Universidad CESMAG, ubicada en la ciudad de Pasto, Colombia, donde será desarrollado y probado en escenarios dinámicos típicos de un entorno universitario.

1 El problema de investigación

1.1 Objeto o tema de investigación

Sistema de navegación en interiores asistido por visión computacional orientado a personas con discapacidad visual.

1.2 Línea de investigación

Sistemas de Automatización y Control. La línea de sistemas de automatización y control de la Universidad CESMAG desarrolla procesos investigativos orientados al modelamiento, simulación, diseño, desarrollo y evaluación de algoritmos de control, sistemas de control, sistemas inteligentes, control de procesos industriales, sistemas embebidos, acondicionamiento y procesamiento de señales, robótica, domótica e inteligencia artificial.

1.3 Sub-línea de Investigación.

La inteligencia artificial es una respuesta al deseo de aproximar el comportamiento humano y el pensamiento racional a diversos sistemas para la solución de determinadas problemáticas a través de diferentes técnicas para la solución de problemas, entre las que se destacan la lógica difusa, las redes neurales, los sistemas neuro-difusos y los algoritmos genéticos. De esta forma los resultados que se desprenden de los procesos investigativos desarrollados en esta línea se orientan a la generación de algoritmos y metodologías que presentan un comportamiento autónomo, dinámico que se adecua a la evolución del entorno.

1.4 Planteamiento o Descripción del Problema

La navegación en interiores plantea desafíos únicos debido a la disposición cambiante de los objetos, el tráfico de personas y las condiciones de iluminación variables. Las soluciones tradicionales, como mapas estáticos y sistemas de posicionamiento, no logran adaptarse de manera óptima a estos cambios ni ofrecen una detección en tiempo real de obstáculos en espacios cerrados.

Esta problemática impacta de forma especialmente marcada en personas con discapacidad visual o baja visión, quienes enfrentan barreras significativas para desplazarse con autonomía dentro de entornos complejos como instalaciones universitarias. Según estimaciones globales, al

menos 2,2 mil millones de personas en el mundo presentan discapacidad visual o ceguera, de las cuales aproximadamente 1 000 millones podrían haber evitado o aún no han recibido tratamiento adecuado [1]. Además, a nivel educativo, menos del 10 % de las personas con discapacidad a nivel mundial acceden a la educación superior, y en regiones como América Latina esta cifra puede variar entre el 2 % y el 5 %, mientras que en Colombia solo alrededor del 0,58 % de los estudiantes matriculados en educación superior son personas con discapacidad; estos datos ponen de manifiesto las profundas brechas de inclusión existentes en el acceso y la permanencia en la educación universitaria para este sector de la población [2].

Por ejemplo, en instalaciones universitarias como bibliotecas, laboratorios y pasillos de facultades, la organización de los espacios puede cambiar con frecuencia. El flujo constante de estudiantes y personal en áreas como escaleras, entradas y salas comunes genera un entorno dinámico y en constante cambio. Estas características hacen que la navegación en interiores, especialmente en edificios con alta densidad de personas y con una estructura compleja, presente retos que las soluciones tradicionales no pueden resolver de forma eficaz.

Un sistema inteligente de navegación asistido por visión computacional, basado en algoritmos avanzados como YOLOv8 y la biblioteca OpenCV, podría ofrecer una solución eficaz en entornos interiores. Este sistema utiliza dos cámaras para capturar imágenes del entorno y emplea estereovisión (con OpenCV) para la estimación de profundidad, además de YOLOv8 para el reconocimiento de objetos. OpenCV también permite detectar colores específicos, facilitando la identificación de señales, objetos y áreas relevantes, lo cual mejora la orientación y seguridad del usuario. Asimismo, el sistema incluye alertas de voz mediante tecnología de *text-to-speech*, utilizando la herramienta Google Text-to-Speech (GTTS), que guía al usuario en tiempo real al informar sobre obstáculos y rutas seguras.

El sistema de navegación asistido por visión computacional se implementará en un computador, donde se ejecutarán los algoritmos de detección y estimación en tiempo real. El sistema utilizará los recursos del hardware del computador para procesar las imágenes y datos capturados. Para facilitar el control y la evaluación del sistema, se desarrollará una interfaz gráfica en el computador, permitiendo monitorear el rendimiento de manera visual durante las pruebas.

Este sistema tiene un gran potencial de aplicación en la navegación en interiores para personas con discapacidad visual o con baja visión, facilitando su desplazamiento seguro en

edificios complejos y en entornos industriales cerrados donde la seguridad es primordial. Al integrarse en dispositivos portátiles, como gafas inteligentes, y proporcionar retroalimentación auditiva, el sistema emitiría alertas sobre obstáculos cercanos y rutas seguras, mejorando la autonomía y seguridad del usuario en espacios interiores.

A pesar de sus ventajas, aún se desconoce la efectividad del sistema de navegación asistido por visión computacional en la prevención de accidentes y en el desplazamiento seguro e independiente de personas con discapacidad visual o baja visión, ya que no se ha realizado un estudio exhaustivo sobre su impacto en la detección y evitación de obstáculos, la identificación de rutas seguras y la orientación en espacios cerrados. Si no se lleva a cabo esta investigación, las posibles aplicaciones del sistema en la navegación en interiores para personas con baja visión y discapacidad visual seguirían limitadas, impidiendo avances significativos en la mejora de su autonomía y seguridad en estos entornos.

1.5 Formulación del Problema

¿Cuál es la efectividad del sistema de navegación en interiores basado en visión computacional orientado a personas con discapacidad visual?

1.6 Objetivos

1.6.1 Objetivo General

Determinar la efectividad del sistema de navegación en interiores asistido por visión computacional orientado a personas con discapacidad visual.

1.6.2 Objetivos Específicos

- Diseñar el sistema de navegación asistido por visión computacional que implemente algoritmos avanzados como YOLOv8, además de estereovisión (con OpenCV), junto con retroalimentación auditiva mediante la herramienta Google Text-to-Speech.

- Implementar el sistema de navegación utilizando cámaras, retroalimentación auditiva y algoritmos de visión computacional para el desplazamiento seguro e independiente de los usuarios, desarrollando una interfaz gráfica en el computador para el monitoreo y control del sistema.

- Evaluar la efectividad del sistema de navegación mediante la observación directa, en un entorno controlado y con personas que simulan las condiciones reales, los investigadores, midiendo el desempeño en términos de precisión, velocidad y costo computacional.

1.7 Justificación

El sistema de navegación asistido por visión computacional propuesto tiene un impacto significativo en la mejora de la autonomía y seguridad de personas con discapacidad visual o baja visión, particularmente en entornos complejos como los de la Universidad CESMAG. La disposición dinámica de los espacios y el flujo constante de personas en áreas como pasillos, escaleras y aulas pueden representar retos adicionales para la orientación de estos usuarios. Este sistema busca facilitar la movilidad de estudiantes y personal con discapacidad visual, ofreciendo una solución de navegación en tiempo real que les permita desplazarse de manera más segura y eficiente dentro del campus.

La implementación de este sistema no solo tiene el potencial de mejorar la calidad de vida y la independencia de las personas con discapacidad visual en entornos interiores, sino que también proporciona una solución avanzada para la detección de obstáculos y peligros en espacios cerrados. Al ofrecer retroalimentación en tiempo real sobre el entorno, el sistema incrementa la seguridad, la confianza y la autonomía de los usuarios, reduciendo significativamente el riesgo de accidentes y ayudando a los usuarios a navegar de manera independiente por entornos como la universidad o cualquier otro espacio complejo.

La tecnología de visión computacional presenta una gran versatilidad, lo que le permite adaptarse a diversas aplicaciones con impacto significativo. Por ejemplo, en instalaciones como centros comerciales, hospitales o edificios universitarios, un sistema de navegación asistido por visión podría facilitar la movilidad de personas con discapacidad visual, ayudándoles a evitar obstáculos como muebles, puertas, escaleras o el tráfico de personas. Además, en situaciones de emergencia, la detección rápida de cambios en el entorno, como objetos temporales o salidas bloqueadas, sería crucial para la seguridad del usuario.

Al proporcionar alertas auditivas y direcciones claras sobre los obstáculos, este sistema permitiría a los usuarios navegar con mayor confianza y realizar actividades que en otros contextos podrían resultar difíciles. Esta solución también destaca por su adaptabilidad a necesidades específicas de distintos grupos. En el ámbito educativo, por ejemplo, este sistema

podría implementarse para apoyar la movilidad de estudiantes en espacios interiores complejos, como edificios universitarios, sin necesidad de involucrar a usuarios finales en esta etapa de desarrollo. El proyecto se enfoca en la creación de un prototipo que, mediante simulaciones y pruebas en escenarios controlados, permite evaluar su efectividad para una posible futura implementación. Esta aproximación permite el desarrollo de un diseño seguro y eficiente, con el potencial de ser adaptado e implementado en aplicaciones prácticas para personas con discapacidad visual por futuros desarrolladores, garantizando el cumplimiento de los principios éticos en su desarrollo.

1.8 Viabilidad

El sistema de navegación en interiores asistido por visión computacional orientado a personas con discapacidad visual es una solución viable y accesible. Se dispone de la información suficiente, así como de las herramientas y el tiempo necesarios para abordar y llevar a cabo el proyecto de investigación de manera eficiente. Además, los recursos financieros se encuentran asegurados mediante fondos propios, eliminando la necesidad de algún tipo de financiamiento externo.

En cuanto a los recursos tecnológicos, el proyecto aprovechará el uso de software gratuito, como Python, y utilizará algoritmos avanzados de código abierto, como el modelo de detección de objetos YOLOv8 y la librería OpenCV para el procesamiento de imágenes y estimación de proximidad.

Para la retroalimentación, el sistema utilizará la versión gratuita de Google Text-to-Speech, la cual ofrece una calidad de voz bastante buena. Dado que el proyecto está diseñado para dar alertas a un solo usuario, y no requiere de una interacción masiva, esta versión gratuita es suficiente para su implementación actual, proporcionando una solución adecuada sin necesidad de servicios premium. Esto asegura que todas las herramientas necesarias para el desarrollo del sistema estén disponibles sin incurrir en costos adicionales, facilitando su implementación de manera accesible y eficiente.

El sistema de navegación se implementará utilizando el hardware de un computador, donde se ejecutarán los algoritmos de visión computacional para la detección y estimación de proximidad en tiempo real. El computador será responsable de procesar las imágenes y datos capturados por las cámaras, sin necesidad de recursos adicionales costosos. Se desarrollará una

interfaz para facilitar el control y la evaluación del sistema. Esta interfaz permitirá monitorear el rendimiento del sistema y realizar ajustes durante las pruebas de manera interactiva, asegurando la facilidad de uso y la efectividad en el proceso de evaluación.

1.9 Delimitación

La implementación del sistema de navegación asistido por visión computacional se llevará a cabo en la Universidad CESMAG, donde se desarrollará y evaluará un prototipo diseñado específicamente para orientar a personas con discapacidad visual o baja visión. El sistema se basará en algoritmos de OpenCV y el modelo de detección de objetos YOLOv8, adaptados para trabajar en entornos interiores, garantizando su efectividad en escenarios dinámicos.

El sistema proporcionará retroalimentación auditiva al usuario mediante Google Text-to-Speech, indicándole de manera precisa y en tiempo real la presencia de obstáculos y rutas seguras, mejorando su movilidad en espacios universitarios. Utilizando Python como el lenguaje principal de programación, se implementarán las funcionalidades necesarias para integrar el modelo de visión computacional con el sistema de navegación, asegurando una respuesta rápida y precisa.

El enfoque principal será garantizar la adaptabilidad del sistema a diferentes condiciones de iluminación y disposición del entorno dentro del campus universitario, permitiendo que el sistema pueda ser evaluado en un entorno controlado antes de una posible futura implementación en otros contextos.

2 Tópicos del Marco Teórico

2.1 Antecedentes

2.1.1 *Detección de objetos en tiempo real con YOLO*

En este artículo [3], se enfatiza en la creación de un sitio web que ayuda a personas con discapacidad visual a identificar objetos en su entorno utilizando el algoritmo YOLOv3. El objetivo principal es facilitar la interacción de personas ciegas con su entorno, proporcionando información sobre los objetos cercanos a través de una retroalimentación auditiva.

No obstante, el estudio se centra principalmente en la viabilidad funcional del sistema y en la correcta detección de objetos, sin realizar un análisis detallado del costo computacional, tales como el uso de recursos, tiempos de inferencia o escalabilidad del modelo en distintos dispositivos. Asimismo, no se exploran comparaciones con versiones más recientes del algoritmo YOLO ni se evalúa su impacto en términos de eficiencia y precisión.

En este contexto, resulta relevante destacar la evolución de las versiones del modelo YOLO para tareas de detección de objetos. En el presente trabajo, se emplea el modelo YOLOv8, el cual corresponde a una versión más reciente y optimizada, ofreciendo mejoras significativas en velocidad de procesamiento, precisión y eficiencia computacional, aspectos clave para sistemas de navegación asistida en tiempo real.

Figura 1

Esquema conceptual de un sistema de detección de objetos con retroalimentación auditiva para personas con discapacidad visual basado en YOLO.

Sistema de Detección para Invidentes



2.1.2 Detección híbrida de objetos para asistencia visual

Este artículo [4], aborda el desarrollo de una tecnología portátil de asistencia (PAT) para personas con discapacidad visual. Propone un sistema que, utilizando un smartphone, detecta y segmenta objetos en el entorno para proporcionar información en tiempo real a los usuarios mediante retroalimentación auditiva. El objetivo es facilitar la comprensión del entorno y ayudar a evitar obstáculos durante la vida diaria de las personas con discapacidad visual.

Utiliza una técnica híbrida que combina el detector de bordes Canny y la Transformada de Hough (HCED-HT) para segmentar objetos capturados por la cámara del smartphone. Así mismo, implementa una versión mejorada del algoritmo YOLOv7, denominada Improved Momentum Search (IMS), que permite una detección más robusta y eficiente de los objetos en el entorno.

De este artículo se rescata la utilidad de una técnica híbrida de segmentación (HCED-HT) y una versión mejorada de YOLOv7 (IMS), reforzando la idea de usar dos algoritmos avanzados en un mismo sistema, como es el caso también del presente documento donde se hará uso de YOLOv8, versión más reciente y poderosa que YOLOv7, en combinación con algoritmos de OpenCV y técnicas de estereovisión.

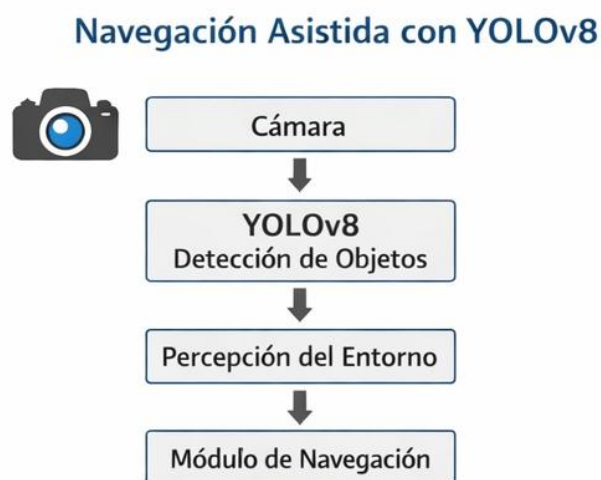
2.1.3 Avances Recientes en Aplicaciones y Métodos de Aprendizaje Profundo para la Navegación Autónoma: Una Revisión Integral

En este artículo [5] se centra en la integración de Deep Reinforcement Learning (DRL) con las capacidades avanzadas de detección de objetos de YOLOv8 dentro de un marco de Simultaneous Localization and Mapping (SLAM). Esta combinación busca mejorar la navegación y la detección de objetos en entornos desafiantes, especialmente interiores, donde variables dinámicas como la iluminación fluctuante, las oclusiones y la presencia de objetos en movimiento representan un reto considerable.

De este estudio resaltamos la capacidad de YOLOv8 para manejar situaciones cambiantes, como iluminación variable y objetos en movimiento, lo que es crucial para un sistema de navegación asistido por visión computacional en entornos interiores. Esto refuerza la importancia de implementar YOLOv8 en el sistema.

Figura 2

Arquitectura general de un sistema de navegación asistida mediante detección de objetos basada en YOLOv8.



2.1.4 Silla de ruedas inteligente para percepción y evitación de obstáculos

En este artículo [6], se presenta el desarrollo de un sistema de Silla de Ruedas Inteligente (Intelligent Wheelchair, IW) diseñado para personas con diversas discapacidades, incluyendo discapacidad visual. Permite al usuario controlar la silla de ruedas mediante movimientos faciales y de la boca, detectados por una cámara. Utiliza cámaras y sensores ultrasónicos para detectar obstáculos y evitar colisiones, como personas o vehículos en movimiento. Clasifica el entorno y determina rutas seguras utilizando un clasificador de texturas basado en máquinas de soporte vectorial (SVM) para navegar de manera segura, evitando colisiones en intersecciones de tráfico.

En el artículo la información del entorno se representa en un mapa de ocupación, donde cada celda refleja el riesgo de la zona correspondiente mediante niveles de gris. Después se utiliza una cámara CCD con una longitud focal de 120° (similar al campo visual humano) y una resolución de 320x240. Finalizando, para crear el mapa de ocupación, se estima un modelo de fondo utilizando aprendizaje en línea simple. El modelo se compara con cada fotograma recibido para clasificarlo en obstáculos o fondo.

De esta investigación es menester destacar la inclusión de un módulo de evaluación de la situación que pueda reconocer diferentes configuraciones de interiores y adaptar la retroalimentación al usuario en función de la complejidad del entorno detectado. La evaluación del método se realizó utilizando imágenes recolectadas en distintos tipos de escenarios.

2.1.5 Caminador inteligente para asistencia de movilidad

Este artículo [7] describe un estudio sobre mejoras tecnológicas en dispositivos de movilidad, como caminadores inteligentes, enfocados en personas con discapacidades tanto de movilidad como visuales. Se diseñaron y estudiaron cuatro configuraciones diferentes de caminadores inteligentes, evaluando diversas tecnologías de detección (como ultrasonido, cámaras de profundidad por infrarrojos y cámaras RGB con procesamiento avanzado de visión por computadora), configuraciones de software y modalidades de interfaz de usuario (como señales hápticas y auditivas).

De este artículo es importante destacar una gran diferencia. Así como el caminador utiliza tecnologías de detección como cámaras y sensores, en el proyecto propuesto a lo largo de este

documento también hace uso de cámaras, tales como el módulo de cámara 3D GXIVISION 4MP, pero OpenCV y la estereovisión ofrecen un enfoque más flexible y poderoso para la percepción del entorno en comparación con los sensores tradicionales, lo que puede ser especialmente beneficioso en aplicaciones de navegación.

2.1.6 Visión Estéreo en Asistencia para la Navegación de Personas Ciegas

En este artículo [8], se describe un proyecto llamado SmartVision, el cual se basa en un sistema modular de navegación diseñado para personas ciegas y con discapacidad visual, que incorpora múltiples tecnologías como GPS, Wi-Fi, RFID, visión por computadora y visión estéreo para proporcionar datos de ubicación en tiempo real, planificación de rutas y evasión de obstáculos. El sistema es versátil, con múltiples redundancias para asegurar su fiabilidad, y proporciona salidas a través de texto a voz.

De este artículo destacamos la importancia de usar la técnica de estéreo visión para estimar la profundidad y proporcionar una navegación segura en tiempo real. Esto ayuda al sistema a calcular las distancias entre el usuario y los objetos detectados. Así mismo, SmartVision nos refuerza la idea de implementar tecnologías de text-to-speech (TTS), proporcionando instrucciones de orientación al usuario a través de texto a voz. Aun así, el artículo no menciona algo relacionado a YOLO, tengamos en cuenta que es un modelo muy versátil al momento de reconocer obstáculos. Y en el proyecto propuesto en este documento se incluye uno de los modelos más avanzados de la familia YOLO, el cual es YOLOv8.

Figura 3

Esquema conceptual del uso de visión estéreo para la estimación de distancia y detección de obstáculos en sistemas de navegación asistida.

Visión Estéreo y Estimación de Distancia



2.1.7 Visión Artificial para Personas con Discapacidad Visual utilizando Detección con YOLO

Este artículo [9] presenta un sistema basado en tecnologías de visión por computadora y procesamiento de imágenes, diseñado para asistir a personas con discapacidad visual en su navegación diaria mediante sustitución de visión. Utiliza el modelo de detección de objetos YOLOv3, conocido por su alta precisión y velocidad, en conjunto con Google Text-to-Speech (GTTS) para convertir los objetos detectados en retroalimentación auditiva en tiempo real. El sistema consta de una webcam para capturar video y auriculares para proporcionar instrucciones al usuario, permitiendo la detección de obstáculos y objetos a mayor rango que herramientas tradicionales como el bastón blanco, con el objetivo de mejorar la independencia y seguridad del usuario en entornos desconocidos o concurridos.

Es importante destacar, al igual que las anteriores investigaciones, el modelo YOLO, el cual viene siendo una excelente herramienta para el reconocimiento de objetos. YOLOv8 representa una mejora significativa sobre YOLOv3 en aspectos clave como velocidad, precisión y facilidad de uso. Gracias a su arquitectura optimizada y técnicas avanzadas de entrenamiento, YOLOv8 ofrece una mayor precisión en la detección de objetos, especialmente en escenarios complejos, mientras mantiene tiempos de inferencia más baja. Así mismo, al igual que el anterior antecedente, este artículo resalta la importancia de un sistema de retroalimentación auditiva,

siendo que los investigadores hacen uso de Google Text-to-Speech, una buena herramienta que utiliza la síntesis de texto a voz.

2.1.8 Conversión de Imágenes a Voz para Personas con Discapacidad Visual

Este artículo [10] aborda el concepto de conversión de imágenes a voz para personas con discapacidad visual, explorando cómo las tecnologías de detección de imágenes y síntesis de voz pueden trabajar juntas para mejorar la accesibilidad y la independencia de las personas con discapacidad visual. El sistema propuesto utiliza OpenCV y YOLO (You Only Look Once) para detectar objetos en imágenes, que luego se transforman en texto, el cual es convertido en voz mediante tecnologías de Text-to-Speech (TTS). La integración de estas tecnologías permite que las personas ciegas o con visión reducida puedan acceder a la información visual a través de medios auditivos, mejorando su capacidad para interactuar con su entorno. El artículo también menciona el uso de redes neuronales y el aprendizaje profundo en la mejora de la precisión de la detección y la calidad de la síntesis de voz.

Aunque el propósito del artículo es interpretar visualmente lo que aparece en una imagen (leer y describir), luego transmitir esa información de manera auditiva usando tecnología de síntesis de voz (TTS), funcionando como una herramienta de lectura y no como un sistema de navegación autónoma como el propuesto en el presente documento, es importante destacar cómo la detección de objetos y la síntesis de voz trabajan juntas. Por lo que la integración de la tecnología text-to-speech (TTS), además de la navegación visual, permitiría a los usuarios recibir una descripción auditiva de los obstáculos y objetos detectados en su entorno, lo que puede enriquecer la experiencia y la precisión en la navegación.

2.1.9 Detección de Objetos en Tiempo Real y Retroalimentación de Audio para Personas con Discapacidad Visual

Este artículo [11] presenta un sistema de detección de objetos en tiempo real y retroalimentación de audio para ayudar a personas con discapacidad visual a identificar y navegar por su entorno. El sistema propuesto utiliza el algoritmo YOLOv3 junto con el conjunto de datos MS COCO para detectar y clasificar objetos, y genera retroalimentación en forma de audio utilizando la API TTS (Google Text-to-Speech). El objetivo es proporcionar una solución

accesible y práctica para mejorar la independencia y accesibilidad de las personas con discapacidad visual.

Aunque YOLOv3 es un modelo robusto para el reconocimiento de objetos, la versión YOLOv8 demuestra ser aún más avanzada y adecuada para aplicaciones de detección en tiempo real. Sin embargo, el artículo mencionado no aborda el uso de la estereovisión, una técnica que puede complementar eficazmente un sistema de navegación, especialmente en la estimación de proximidad. En el proyecto propuesto en este documento, se combinarán las ventajas de la estereovisión para mejorar la estimación de distancias con la potencia y versatilidad de YOLOv8, lo que potenciará la precisión y fiabilidad del sistema de navegación asistido.

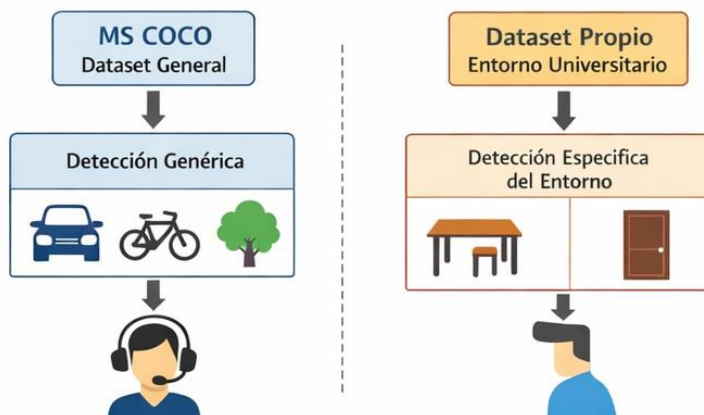
Además, el artículo hace referencia a un conjunto de datos MS COCO, que es el conjunto con la que generalmente viene pre-entrenado YOLO. Aunque MS COCO es un excelente conjunto de datos (dataset) para tareas generales de detección, aun así, dado que el sistema de navegación en el contexto de este proyecto será evaluado en las instalaciones de la Universidad CESMAG, un dataset tan general como MS COCO es útil para la detección de personas, pero no es completamente adecuada para el contexto específico de las instalaciones de la universidad. Esto se debe a que la base está diseñada para reconocer una amplia variedad de objetos en entornos más generales, como calles y espacios públicos, mientras que las instalaciones universitarias pueden presentar características particulares. Afortunadamente, YOLOv8 tiene la capacidad de ser entrenado con nuevos conjuntos de datos, lo que permite adaptar el modelo para detectar una gama más amplia de objetos específicos relevantes para el entorno local y las necesidades del proyecto.

Finalmente, es crucial destacar nuevamente la tecnología TTS (Text-to-Speech), que en el artículo se implementa utilizando Google Text-to-Speech. Esta herramienta resulta ser muy útil, ya que emplea tecnología avanzada de síntesis de voz para convertir texto escrito en audio, lo que facilita la interacción con personas con discapacidad visual.

Figura 4

Comparación conceptual entre el uso de un conjunto de datos general y un conjunto de datos específico del entorno para la detección de objetos.

Conjuntos de Datos: MS COCO vs. Dataset Propio



2.1.10 Third Eye: Reconocimiento de Objetos y Generación de Habla para Personas con Discapacidad Visual

Este artículo [12] describe un sistema para ayudar a las personas con discapacidad visual a identificar y comprender mejor su entorno mediante la detección y reconocimiento de objetos, generando descripciones de los mismos a través de habla. El sistema propuesto, denominado "Third Eye", utiliza el modelo de detección de imágenes YOLOv5 junto con conversores de texto a voz como los módulos gTTS y pyttsx3 en Python. El objetivo es permitir que las personas ciegas o con discapacidad visual viajen de manera independiente, identificando los objetos en su camino y generando descripciones de estos objetos detectados. YOLOv5 se entrena con un conjunto de datos personalizado de 15 objetos, junto con el conjunto de datos MS COCO 2017 que contiene 80 objetos (en total, 95 objetos). Las etiquetas de salida del modelo se transforman en texto, que luego se convierte a formato de audio y se presenta al usuario a través de un altavoz. Además, se comparan dos bibliotecas de Python para la conversión de audio: pyttsx3 y gTTS.

De este artículo es importante destacar la existencia de diversas herramientas para implementar tecnología de conversión de texto a voz (Text-to-Speech, TTS). No solo Google TTS es una opción viable, sino que también existen alternativas como la biblioteca pyttsx3, que utiliza motores de voz locales. Dependiendo de su rendimiento, pyttsx3 podría ser considerada para su integración en el sistema propuesto. Así mismo, al igual que otros antecedentes mencionados anteriormente, este artículo también hace referencia a la base de datos MS COCO,

que es ampliamente utilizada en el entrenamiento de modelos YOLO. Aunque esta base de datos está más orientada a la detección de objetos en entornos exteriores que en interiores, su capacidad para detectar personas es valiosa y relevante para el sistema que se propone desarrollar.

Tabla 1

Principal aporte de investigación

Antecedente	Versión YOLO	Esterovisión	Estimación de distancia	Retroalimentación auditiva	Dataset personalizado	Asignación zona espacial	Evaluación end-to-end	Métricas computacionales reportadas	Entorno interior específico
[3]	YOLOv3	No	No reportada	Sí (audio)	No (COCO)	No	No	No reportadas	No (sitio web)
[4]	YOLOv7 mejorado	No	No reportada	Sí (audio)	No	No	No	No reportadas	No especificado
[5]	YOLOv8 (mención teórica)	No	Mencionada, sin margen de error	No	No	No	No	Parcialmente	Sí (entre otros)
[6]	No usa	No	Sí (sensores ultrasónicos)	No	No	No	Sí (silla de ruedas)	No reportadas	Sí
[7]	No usa	No	Sí (cámaras IR y ultrasonido)	Sí (háptica y audio)	No	No	No	No reportadas	Sí
[8]	No usa	Sí	Sí (sin margen de error definido)	Sí (TTS)	No	No	No	No reportadas	No especificado
[9]	YOLOv3	No	No reportada	Sí (audio)	No (COCO)	No	No	No reportadas	No
[10]	YOLO (no especificada)	No	No reportada	Sí (TTS)	No	No	No	No reportadas	No
[11]	YOLOv3	No	No reportada	Sí (gTTS)	No (COCO)	No	No	No reportadas	No
[12]	YOLOv5	No	No reportada	Sí (gTTS y pyttsx3)	Sí (15 clases)	No	No	No reportadas	No especificado
Trabajo presente	YOLOv8n	Sí (OpenCV)	Sí (± 10 cm, documentado)	Sí (gTTS)	Sí (29 clases, entorno universitario)	Sí (7 zonas)	Sí (60 recorridos, 3 escenarios)	Sí (FPS, CPU, GPU, RAM)	Sí (Universidad CESMAG)

El análisis comparativo presentado en la tabla anterior permite identificar con claridad el valor diferencial de esta investigación respecto a los trabajos previos revisados. En primer lugar, ninguno de los antecedentes integra simultáneamente los cuatro componentes centrales del sistema propuesto: detección y clasificación de objetos mediante YOLOv8, estimación de distancias por estereovisión con margen de error documentado, asignación de zona espacial y retroalimentación auditiva en tiempo real. Cada antecedente aborda uno o dos de estos componentes de forma aislada, mientras que el presente trabajo los integra como un sistema funcional unificado orientado a la navegación asistida en entornos interiores.

En segundo lugar, el presente trabajo es el único que reporta un margen de error definido y validado experimentalmente para la estimación de distancias (± 10 cm), mientras que los antecedentes que abordan esta funcionalidad, como [8], [6] y [7], no establecen un criterio de tolerancia cuantificable ni lo validan mediante pruebas repetidas.

En tercer lugar, el sistema incorpora un módulo de asignación de zona espacial con siete regiones del campo visual, ausente en todos los antecedentes revisados, lo que permite al usuario recibir información no solo sobre la presencia y distancia de un obstáculo, sino también sobre su ubicación relativa en el entorno.

En cuarto lugar, a diferencia de la mayoría de los antecedentes que utilizan datasets genéricos como MS COCO, el presente trabajo desarrolló un dataset personalizado de 28 clases específicas para el entorno universitario interior de la Universidad CESMAG, superando en especialización contextual al trabajo más cercano en este aspecto, que es el [12] con 15 clases personalizadas.

Finalmente, el presente trabajo evalúa el sistema de forma end-to-end mediante 60 recorridos en 3 escenarios controlados, lo que permite caracterizar el desempeño global del sistema de manera integral y reproducible.

2.2 Enunciados de los supuestos teóricos

2.2.1 Visión computacional

La visión se basa en reconocer y localizar objetos en el ambiente mediante el procesamiento de las imágenes. La visión computacional es el estudio de estos procesos, para entenderlos y construir máquinas con capacidades similares.

La visión computacional está estrechamente relacionada con el procesamiento de imágenes, aunque sus objetivos son diferentes. El procesamiento de imágenes se enfoca en mejorar la calidad visual (remover defectos, desenfoque, mejorar colores y contrastes), mientras que la visión computacional busca extraer características de las imágenes para su análisis e interpretación automática, como localizar y clasificar objetos [13]. En este documento los algoritmos a usar son:

OpenCV (Open Source Computer Vision): Es una biblioteca de código abierto diseñada para el procesamiento de imágenes en tiempo real. Desarrollada inicialmente por Intel en el año 2000, ha crecido hasta convertirse en una herramienta estándar en visión computacional, con soporte para múltiples plataformas. Originalmente implementada en C, su versión 2.0 añadió soporte para C++ y más recientemente incluye interfaces para lenguajes como Python, Java y MATLAB. OpenCV ofrece funciones optimizadas para el procesamiento de imágenes y video. Sus aplicaciones van desde reconocimiento de objetos y rostros, hasta seguimiento de movimientos y realidad aumentada, integrando además herramientas de aprendizaje automático para mejorar el rendimiento en tareas complejas de visión artificial. [14]

2.2.2 *Redes neuronales convolucionales (CNN)*

Las Redes Neuronales Convolucionales (CNN) constituyen una de las arquitecturas más relevantes dentro del aprendizaje profundo, especialmente diseñadas para el análisis de datos con estructura espacial, como imágenes y secuencias de video. Su principal fortaleza radica en la capacidad de aprender representaciones jerárquicas directamente a partir de los datos, permitiendo identificar patrones desde niveles básicos hasta conceptos semánticos complejos sin necesidad de diseñar manualmente descriptores [15].

A diferencia de los enfoques tradicionales de visión artificial, donde las características debían ser definidas explícitamente (por ejemplo, bordes o esquinas), las CNN realizan este proceso de forma automática mediante un aprendizaje supervisado. Esto les permite adaptarse mejor a variaciones en iluminación, escala, orientación y oclusión, aspectos comunes en entornos reales.

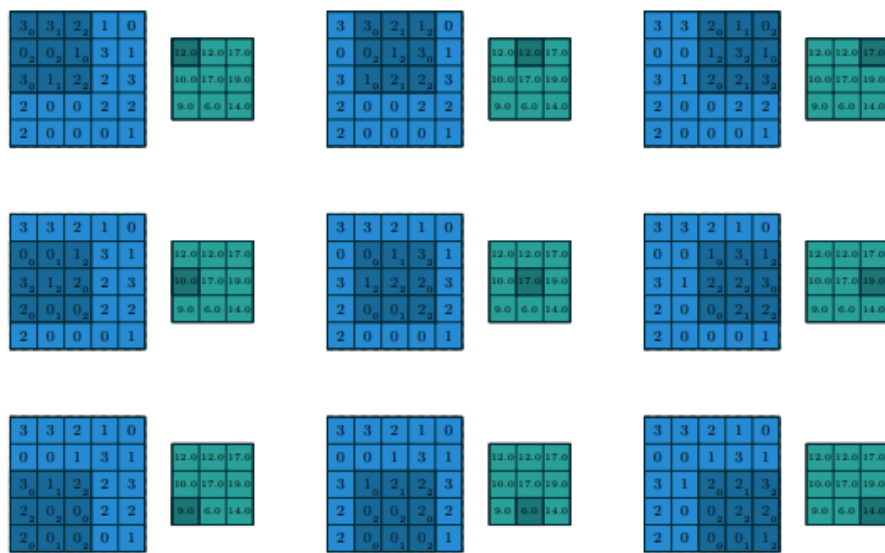
La estructura de una CNN está compuesta por múltiples capas organizadas de manera secuencial, cada una encargada de transformar progresivamente la información de entrada en representaciones más abstractas:

Capas convolucionales:

Constituyen el núcleo de la red. En estas capas, se aplican filtros (kernels) que se deslizan sobre la imagen de entrada realizando operaciones de convolución. Este proceso genera mapas de características que resaltan patrones locales como bordes, texturas o formas. A medida que se incrementa la profundidad de la red, estos patrones evolucionan hacia representaciones más complejas, como partes de objetos o estructuras completas.

Figura 5

Convolución discreta en 2D — Dumoulin y Visin [15]



Capas de activación:

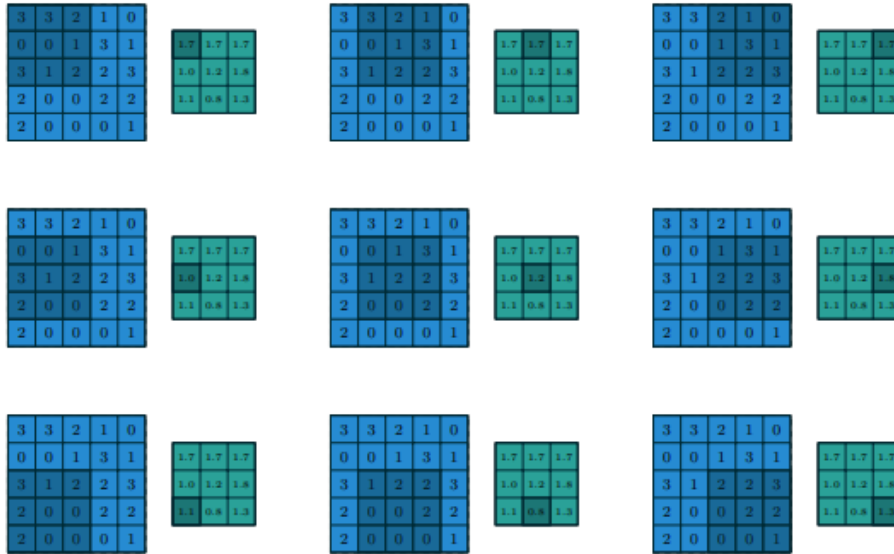
Introducen no linealidad en el modelo, permitiendo que la red represente relaciones complejas entre los datos. La función más utilizada es la ReLU (Rectified Linear Unit), la cual activa únicamente valores positivos, facilitando el entrenamiento y mitigando problemas como el desvanecimiento del gradiente.

Capas de agrupamiento (pooling):

Reducen la dimensionalidad de los mapas de características, conservando únicamente la información más relevante. Este proceso no solo disminuye la carga computacional, sino que también aporta invariancia frente a pequeñas transformaciones en la imagen, como traslaciones o deformaciones leves.

Figura 6

Average Pooling 5×5 — Dumoulin y Visin [15]



Capas completamente conectadas (fully connected):

Ubicadas generalmente en la parte final de la red, integran todas las características extraídas previamente para generar la salida del modelo. Dependiendo de la tarea, esta salida puede corresponder a una clasificación, una regresión o parámetros asociados a la detección de objetos.

En conjunto, esta arquitectura permite que las CNN sean altamente efectivas en tareas de visión por computador, constituyendo la base de modelos modernos como los detectores de objetos en tiempo real.

2.2.3 Algoritmos de Deep Learning para visión computacional.

El algoritmo YOLO (You Only Look Once), propuesto por Redmon et al. en 2016, introdujo un cambio significativo en el paradigma de detección de objetos al plantear un enfoque unificado y eficiente para esta tarea [16]. A diferencia de los métodos tradicionales que separaban el proceso en múltiples etapas (como generación de regiones candidatas y posterior clasificación), YOLO reformula la detección como un único problema de regresión directa.

En este enfoque, una red neuronal convolucional procesa la imagen completa en una sola pasada, prediciendo simultáneamente las ubicaciones de los objetos y sus respectivas clases. Esto permite reducir considerablemente el tiempo de inferencia, haciendo viable su uso en aplicaciones en tiempo real.

El funcionamiento del modelo se basa en dividir la imagen de entrada en una cuadrícula de tamaño $S \times S$. Cada celda de la cuadrícula es responsable de predecir:

- B cuadros delimitadores (bounding boxes)
- Un valor de confianza para cada cuadro
- Probabilidades de pertenencia a cada una de las C clases

La salida total del modelo puede expresarse como:

$$Salida = S \cdot S \cdot (B \cdot 5 + C) \quad (1)$$

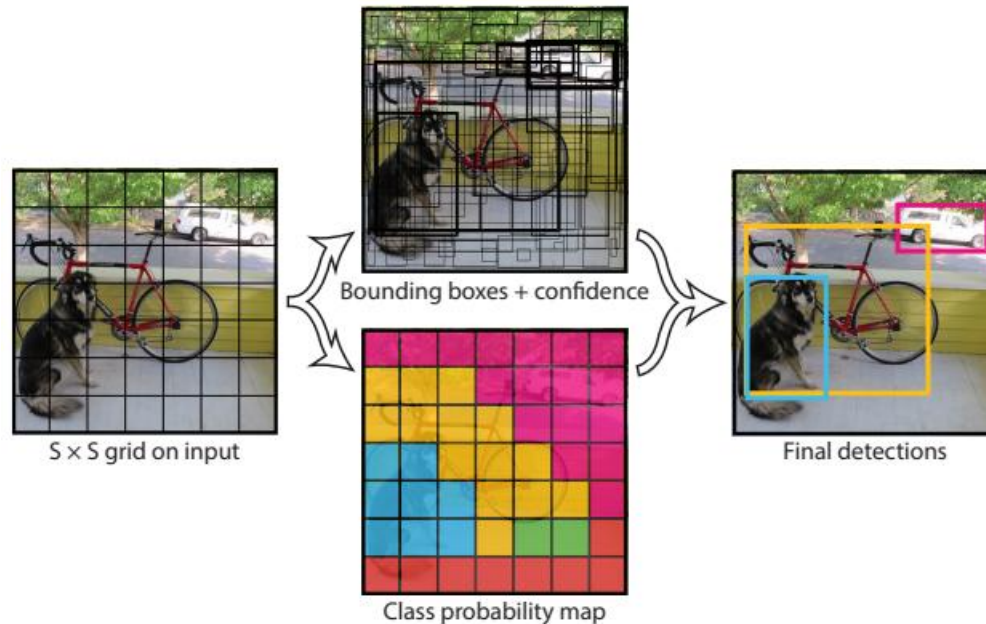
Donde:

- $S \cdot S$: número de celdas en la imagen
- B : número de bounding boxes por celda
- 5: parámetros de cada bounding box ($x, y, w, h, Confianza$)
- C : número de clases

Este diseño permite que YOLO capture tanto información espacial como contextual en una única inferencia, logrando velocidades de procesamiento del orden de milisegundos. Sin embargo, una de sus limitaciones iniciales radicaba en la dificultad para detectar objetos pequeños o múltiples objetos cercanos dentro de una misma celda.

Figura 7

División en cuadrícula y predicción de bounding boxes — Redmon et al. [16]



Arquitectura de YOLOv8

A partir del modelo original, la familia YOLO ha evolucionado progresivamente incorporando mejoras en precisión, velocidad y estabilidad. Versiones intermedias como YOLOv2, YOLOv3, YOLOv4 y YOLOv5 introdujeron avances en la profundidad de la red, estrategias de entrenamiento y optimización de funciones de pérdida.

YOLOv8, desarrollado por Ultralytics, representa una evolución reciente que incorpora cambios arquitectónicos importantes orientados a mejorar la eficiencia y simplificar el proceso de entrenamiento, manteniendo un alto desempeño en tareas de detección en tiempo real [17].

La arquitectura de YOLOv8 se organiza en tres bloques principales:

1. Backbone (extracción de características):

Encargado de procesar la imagen de entrada y generar representaciones jerárquicas. YOLOv8 utiliza una variante optimizada basada en principios de CSPNet, incorporando el módulo C2f, el cual mejora el flujo de información entre capas y permite una extracción de características más eficiente con menor costo computacional.

2. Neck (fusión multi-escala):

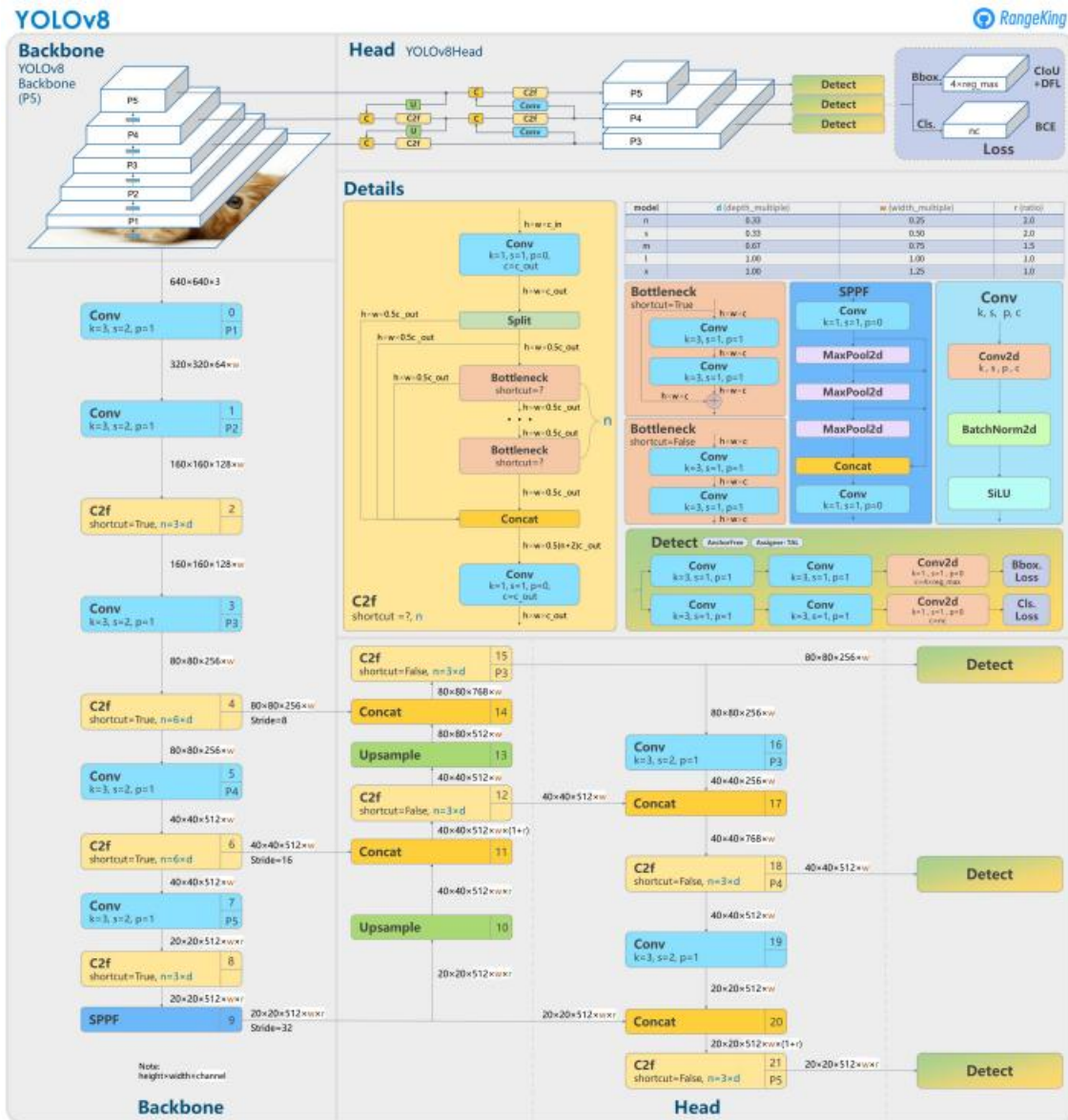
Este bloque integra estructuras como Feature Pyramid Network (FPN) y Path Aggregation Network (PAN), permitiendo combinar información proveniente de diferentes niveles de la red. Gracias a esta fusión multi-escala, el modelo puede detectar objetos de distintos tamaños con mayor precisión.

3. Head (predicción):

YOLOv8 adopta un enfoque anchor-free, eliminando la necesidad de cajas ancla predefinidas utilizadas en versiones anteriores. En su lugar, predice directamente las coordenadas de los objetos, lo que reduce la complejidad del modelo y mejora su capacidad de generalización frente a diferentes formas y proporciones.

Figura 8

Arquitectura completa YOLOv8 — Yaseen [17]



Adicionalmente, YOLOv8 incorpora mejoras en el proceso de entrenamiento, como técnicas avanzadas de aumentación de datos (por ejemplo, mosaico y combinaciones de imágenes), así como funciones de pérdida optimizadas que permiten un mejor ajuste tanto en la clasificación como en la localización de objetos.

En conjunto, estas características posicionan a YOLOv8 como una solución robusta y eficiente para aplicaciones de visión computacional en tiempo real, siendo especialmente

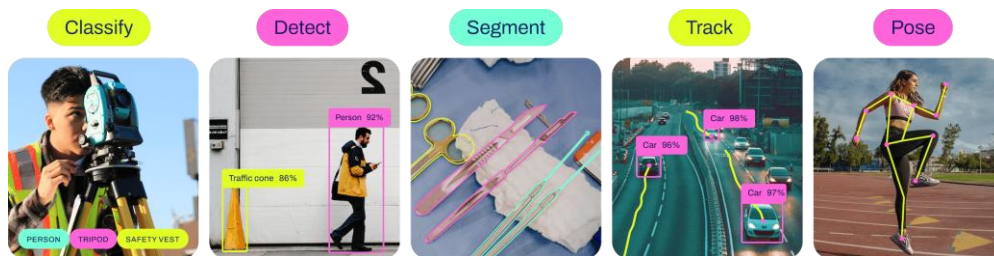
adecuado para sistemas embebidos y entornos donde la latencia y el rendimiento son factores críticos.

Ecosistema de código abierto y comunidad

YOLOv8 se distribuye bajo licencia AGPL-3.0 para uso educativo y no comercial, y bajo una licencia Enterprise para aplicaciones comerciales. La licencia AGPL-3.0, aprobada por OSI, promueve la colaboración abierta y el intercambio de conocimiento, garantizando que cualquier mejora derivada del modelo sea devuelta a la comunidad. El modelo está disponible públicamente en el repositorio oficial de Ultralytics en GitHub, junto con documentación detallada, ejemplos de uso y modelos pre-entrenados para diversas tareas de visión computacional. YOLOv8 ofrece una variedad de modelos pre-entrenados orientados a distintos requisitos de rendimiento, lo que facilita su adopción en diferentes escenarios de aplicación. Esta disponibilidad abierta ha favorecido su adopción en proyectos académicos e industriales, incluyendo sistemas de navegación asistida basados en visión computacional como el desarrollado en esta investigación [18].

Figura 9

Modelo YOLOv8 en la detección, segmentación, clasificación y estimación de postura.[19]



2.2.4 Estereovisión

La estereovisión es una técnica dentro de la visión computacional que permite estimar la profundidad de un entorno tridimensional utilizando dos cámaras colocadas horizontalmente (una a la izquierda y otra a la derecha). Ambas cámaras capturan imágenes simultáneamente, y luego se procesan para recuperar la información de profundidad visual. El desafío principal es generar un mapa de disparidad, que corresponde a las diferencias entre los píxeles de ambas imágenes, permitiendo así reconstruir las coordenadas tridimensionales de los puntos en el espacio. [20]

2.2.5 Discapacidad visual

La discapacidad visual se refiere a la pérdida parcial o total de la capacidad para ver. Esta condición puede abarcar una variedad de problemas, desde visión borrosa y dificultad para distinguir colores, hasta ceguera completa. Las causas pueden incluir enfermedades hereditarias, lesiones o condiciones médicas como la diabetes o el glaucoma. La discapacidad visual afecta la calidad de vida, limitando la movilidad, la independencia y la participación en actividades diarias. Las personas con discapacidad visual pueden utilizar tecnologías asistidas, como lectores de pantalla y dispositivos de navegación, para mejorar su accesibilidad. La rehabilitación y la educación también son fundamentales para ayudarles a adaptarse y superar barreras. [21]

2.2.6 Sistema de navegación

El sistema de navegación se refiere al conjunto de herramientas y tecnologías utilizadas para guiar a una persona en tiempo real, especialmente en la detección y evitación de obstáculos. Este sistema combina cámaras, algoritmos de procesamiento de imágenes (como YOLO para detección de objetos) y técnicas de estereovisión para medir distancias y reconocer el entorno, brindando asistencia en la orientación espacial para facilitar la movilidad segura en interiores.

2.2.7 Salud visual

El concepto de salud ha evolucionado hacia una visión más amplia, integrando aspectos físicos, mentales y sociales, y superando la antigua noción de salud como simple ausencia de enfermedad. La salud visual podría entenderse como un estado integral que va más allá de la ausencia de problemas visuales, incluyendo factores de bienestar y calidad de vida asociados a la visión en su contexto social y ambiental.[22]

2.2.8 Entornos interiores

Entornos interiores se refiere a los espacios dentro de edificios y estructuras cerradas, como oficinas, centros comerciales, hospitales y estaciones de transporte, en los cuales es común el uso de sistemas de localización y navegación específicos. Estos sistemas, al no depender de tecnologías satelitales como el GPS debido a su limitada precisión en interiores, aprovechan

redes inalámbricas, Bluetooth y campos magnéticos para proporcionar datos de ubicación precisos dentro de estos espacios. Los entornos interiores son, por tanto, áreas delimitadas donde se emplean tecnologías especializadas para realizar tareas como el seguimiento de activos, la asistencia en la navegación de personas, y la optimización de operaciones en logística y seguridad. [23]

2.2.9 *Sistemas de procesamientos de datos*

El procesamiento de datos es un proceso que implica la recolección de datos primarios, los cuales son evaluados y organizados para generar información valiosa. Esta información será luego analizada por el usuario, quien podrá tomar decisiones o realizar acciones en función de los resultados obtenidos. En esencia, el procesamiento de datos se refiere a la transformación de datos crudos en información útil y significativa.

En cualquier sistema de procesamiento de datos (SPD), este ciclo consta de seis etapas fundamentales:

Origen: donde los datos se generan o recogen desde diversas fuentes.

Entrada: los datos ingresan al sistema para su posterior procesamiento.

Proceso: los datos se manipulan y se analizan según las necesidades del sistema.

Salida: la información procesada se presenta de manera comprensible para el usuario.

Distribución/Impresión: la información puede ser distribuida o impresa para su uso posterior.

Almacenamiento/Base de datos: los datos y la información resultante se almacenan en una base de datos para su acceso futuro. [24]

2.2.10 *Baja visión*

La baja visión se refiere a un trastorno visual que limita la capacidad para realizar actividades cotidianas que requieren visión. Este tipo de deficiencia no puede ser corregido con gafas comunes, lentes de contacto ni con tratamientos médicos. Entre las manifestaciones más comunes se incluyen la disminución de la agudeza visual y la pérdida de campo visual. Además, otras condiciones asociadas son la reducción en la sensibilidad al contraste, dificultades con la visión del color y la visión nocturna, o un aumento en la sensibilidad a la luz, como la fotofobia o la dificultad para soportar el deslumbramiento. [25]

2.2.11 Retroalimentación auditiva

La retroalimentación auditiva hace referencia a los sonidos que percibimos después de realizar una acción. Por ejemplo, si un lápiz cae al suelo, el sonido que escuchamos es la retroalimentación auditiva resultante de esa acción. Esta retroalimentación es una parte integral de todas las actividades que realizamos, como escribir, cerrar una puerta, colgar el teléfono, abrir un grifo o hablar. Es fundamental para la producción del habla, especialmente cuando los niños pequeños comienzan a aprender a hablar. Ellos escuchan a las personas a su alrededor y tratan de imitar los sonidos que oyen, ajustando sus propias emisiones sonoras basándose en la retroalimentación auditiva que reciben (los sonidos que ellos mismos producen). A medida que los niños crecen y se desarrollan, esta retroalimentación sigue siendo esencial para el aprendizaje.[26]

Text-to-Speech: El Text-to-Speech permite a los desarrolladores crear voces humanas sintéticas con un sonido natural como audio reproducible. Puedes usar los archivos de datos de audio creados con Text-to-Speech para potenciar tus aplicaciones o mejorar los medios como videos o grabaciones de audio (de acuerdo con las Condiciones del Servicio de Google Cloud Platform, incluido el cumplimiento de las leyes aplicables).

Text-to-Speech convierte textos o entradas de lenguaje de marcación de síntesis de voz (SSML) en datos de audio como MP3 o LINEAR16 (la codificación utilizada en archivos WAV). [27]

Google Text-to-Speech (GTTS) es una tecnología de síntesis de voz desarrollada por Google que convierte texto escrito en audio. [27]

2.2.12 Modelos pre entrenados

Un modelo pre entrenado en aprendizaje profundo es un modelo de inteligencia artificial que ha sido previamente entrenado en un conjunto de datos amplio y generalizado, adquiriendo conocimientos que pueden ser transferidos a tareas específicas. Este enfoque se basa en la transferencia de aprendizaje, donde el modelo, que ya ha aprendido a reconocer patrones y características relevantes en datos masivos, puede adaptarse eficientemente a una nueva tarea relacionada, incluso si los conjuntos de datos utilizados son diferentes. Al utilizar modelos pre

entrenados, se reduce el tiempo y los recursos necesarios para entrenar un modelo desde cero, mejorando así la eficiencia y efectividad en la obtención de resultados en tareas específicas. [28]

2.2.13 Bounding Box

Bounding box (caja delimitadora), es un rectángulo alineado a los ejes que envuelve completamente un objeto detectado en una imagen, definido por sus coordenadas extremas (x_1 , y_1 , x_2 , y_2) o por su centro, ancho y alto. Es la representación estándar de salida en algoritmos de detección de objetos [29].

2.2.14 Sistema completo (end-to-end)

En el contexto de este trabajo, "evaluación end-to-end" se refiere al análisis del desempeño del sistema completo desde la captura de imagen hasta la navegación exitosa del usuario, integrando todos los módulos del pipeline (detección, clasificación, estimación de distancia y asignación espacial) en condiciones operativas reales. Esta evaluación difiere del análisis por módulo individual, al considerar el efecto acumulativo de errores en la cadena de procesamiento sobre el resultado funcional final [30].

2.3 Definición de conceptos

2.3.1 Definición nominal de conceptos

La variable principal a medir es la efectividad (Ef) del sistema de navegación asistido por visión computacional, entendida como la capacidad global del prototipo para asistir al usuario durante la ejecución de recorridos en entornos interiores controlados, favoreciendo la detección de obstáculos, la correcta interpretación espacial del entorno inmediato y la generación oportuna de instrucciones auditivas que contribuyan a la finalización segura del trayecto.

En este contexto, la efectividad integra dos dimensiones complementarias: la eficacia operativa, asociada a la capacidad del sistema para permitir la culminación exitosa de los recorridos asistidos, y la eficiencia funcional, relacionada con el balance entre el costo temporal introducido por el sistema y la mejora en seguridad durante la navegación, medida a partir de la reducción de colisiones.

Adicionalmente, la evaluación del sistema contempla métricas de desempeño técnico por componente, tales como precisión (*precision*), recall, F1-score, accuracy, velocidad de procesamiento y costo computacional, las cuales permiten caracterizar cuantitativamente el comportamiento interno del prototipo como variables complementarias de análisis.

2.3.2 *Definición operativa de conceptos*

Para efectos de evaluación experimental, la efectividad (Ef) del sistema se define como la integración ponderada entre dos dimensiones: eficacia y eficiencia.

La eficacia se cuantifica como la proporción de recorridos completados exitosamente sin colisiones (N_e) respecto al total de recorridos ejecutados (N_p) en condiciones controladas, expresada en porcentaje. Para su cálculo se adopta una expresión análoga a la utilizada en sistemas de asistencia similares [31]

$$Ec = \frac{N_e}{N_p} \cdot 100\% \quad (2)$$

donde un recorrido se considera exitoso cuando el usuario logra completar el trayecto recibiendo instrucciones correctas y oportunas por parte del sistema, sin presentar colisiones con obstáculos presentes en el entorno.

La eficiencia se estima a partir de dos indicadores complementarios: un indicador temporal (S_t), que mide la relación entre el tiempo promedio de recorrido sin asistencia y con asistencia, y un indicador de seguridad (S_s), que cuantifica la reducción relativa del número de colisiones con el uso del sistema:

$$S_t = \frac{T_{sin}}{T_{con}} \quad (3)$$

$$S_s = 1 - \frac{C_{con}}{C_{sin}} \quad (4)$$

Ambos indicadores se integran mediante una ponderación diferencial, asignando un 30% de importancia al desempeño temporal y un 70% a la seguridad en navegación:

$$E_f = 0.3S_t + 0.7S_s \quad (5)$$

Finalmente, la efectividad global del sistema se obtiene mediante la integración ponderada de la eficacia y la eficiencia, asignando un peso del 80% a la eficacia y del 20% a la eficiencia:

$$Ef = 0.8E_c + 0.2E_f \quad (6)$$

De manera complementaria, el desempeño técnico de los módulos internos se caracteriza mediante métricas como precision, recall, F1-score y accuracy, utilizadas para analizar el comportamiento individual de los componentes del sistema, sin constituir directamente la medida de efectividad global.

La precisión (P) se refiere a la capacidad del sistema para identificar correctamente los obstáculos en el entorno. Para medir la precisión, se toma como referencia a [32]

$$P = \frac{TP}{TP + FP} \quad (7)$$

En este trabajo

$$P = \frac{TP}{TP + EP} \quad (8)$$

Recall (R) mide la proporción de casos positivos reales que fueron correctamente identificados. Indica qué tan completo es el sistema en detectar todos los objetos presentes [33].

$$R = \frac{TP}{TP + FN} \quad (9)$$

En evaluación end-to-end

$$Recall = \frac{TP}{TP + EP + FN} \quad (10)$$

Accuracy (A) mide la proporción de predicciones correctas sobre el total de casos evaluados. Representa el desempeño global del sistema [34].

$$Accuracy = \frac{TP}{TP + FP} \quad (11)$$

F1-Score (F1) es la media armónica entre Precision y Recall, proporcionando una métrica balanceada que penaliza desbalances extremos entre ambas [35].

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (12)$$

La velocidad (V) mide el tiempo que tarda el sistema en procesar la información y responder a los cambios en el entorno. Se puede expresar en milisegundos por fotograma (ms/frame) y se puede calcular promediando el tiempo de respuesta de cada detección.

Donde

Ef: Efectividad global del sistema.

Ef: Eficiencia del sistema.

Ec: Eficacia del sistema.

St: Indicador temporal.

Ss: Indicador de seguridad.

Tsin: Tiempo promedio de recorrido sin asistencia del sistema

Tcon: Tiempo promedio de recorrido con asistencia del sistema.

Csin: Número promedio de colisiones sin asistencia del sistema.

Ccon: Número promedio de colisiones con asistencia del sistema.

P: Precisión (Precision).

F1: F1-Score

A: Accuracy

R: Recall

N_e : es el número de recorridos de prueba en los que el sistema operó exitosamente.

N_p : es el número total de recorridos de prueba realizados.

TP: (True Positives) son objetos detectados, clasificados correctamente, con distancia dentro de ± 10 cm y zona correcta.

FP: (False Positives) son predicciones incorrectas en módulos individuales

EP: (Error Parcial) son objetos detectados con al menos un criterio incorrecto (clasificación, distancia o zona)

FN: (False Negatives) son objetos presentes no detectados por el sistema.

V: Velocidad.

Basándonos en [36], podemos medir el costo computacional de manera eficiente descomponiéndolo en varias métricas clave relacionadas con el rendimiento del sistema, como el uso de la CPU (medido en porcentaje o megahertz), GPU (en porcentaje o vatios), RAM (medido en megabytes, gigabytes o porcentaje), y otros recursos como el tiempo (segundos) y la temperatura (medida en grados Celsius). De esta manera se puede definir el costo computacional del sistema, donde cada una de las métricas afecta el rendimiento general.

2.4 Hipótesis

2.4.1 Hipótesis de investigación

El sistema de navegación asistido por visión computacional, que utiliza el algoritmo YOLOv8 junto con técnicas de estereovisión para el reconocimiento de objetos y la estimación de profundidad en entornos interiores, presenta una efectividad del 85%. Esta efectividad se evalúa en términos de precisión, velocidad de procesamiento y costo computacional, lo que asegura su rendimiento adecuado en tiempo real dentro de un entorno controlado.

2.4.2 *Hipótesis nula*

El sistema de navegación asistido por visión computacional, que utiliza el algoritmo YOLOv8 junto con técnicas de estereovisión para el reconocimiento de objetos y la estimación de profundidad en entornos interiores, no presenta una efectividad del 85%. Esta efectividad, evaluada en términos de precisión, velocidad de procesamiento y costo computacional, no asegura un rendimiento adecuado en tiempo real dentro de un entorno controlado.

2.4.3 *Hipótesis alternativa*

El sistema de navegación asistido por visión computacional, que utiliza el algoritmo YOLOv8 junto con técnicas de estereovisión para el reconocimiento de objetos y la estimación de profundidad en entornos interiores, presenta una efectividad del 75%. Esta efectividad se evalúa en términos de precisión, velocidad de procesamiento y costo computacional, lo que asegura su rendimiento adecuado en tiempo real dentro de un entorno controlado.

3 Metodología

3.1 Enfoque

El enfoque de esta investigación es cuantitativo [34], ya que se evalúa de manera numérica las variables asociadas al desempeño del sistema de navegación asistido por visión computacional, basado en el modelo YOLOv8 y técnicas de estereovisión con OpenCV. Se mide el comportamiento del sistema a través de experimentos controlados, centrados en métricas objetivas relacionadas con la percepción del entorno mediante visión por computador, la clasificación, la estimación de distancia, la asignación espacial, la velocidad de procesamiento y el costo computacional. Los resultados obtenidos, representados en términos porcentuales y métricas de rendimiento, permite evaluar el desempeño del sistema en escenarios controlados.

Cabe resaltar que la evaluación realizada corresponde a una validación técnica del sistema en etapa de prototipo, desarrollada en entornos controlados sin participación de personas con discapacidad visual. Esta decisión responde a dos criterios fundamentales. En primer lugar, desde una perspectiva ética, la evaluación de un sistema prototipo con personas pertenecientes a poblaciones vulnerables requiere la aprobación previa de un comité de ética institucional, dado que un fallo del sistema durante las pruebas podría comprometer la integridad física de los participantes. Este criterio fue además respaldado por el juicio de expertos en ingeniería consultados durante el desarrollo del proyecto, quienes recomendaron que, en etapas tempranas de prototipado, los propios investigadores simulen las condiciones de uso con el fin de evitar riesgos innecesarios sobre poblaciones vulnerables. En segundo lugar, desde una perspectiva técnica, en el desarrollo de sistemas de visión por computador aplicados a dispositivos de asistencia es práctica común realizar una fase inicial de validación técnica en entornos controlados, con el objetivo de verificar el desempeño de los módulos computacionales y garantizar la estabilidad del sistema antes de proceder a evaluaciones con usuarios finales.

3.2 Paradigma

Desde el punto de vista epistemológico y metodológico, el paradigma que orienta esta investigación es el neopositivismo, una corriente filosófica que busca estudiar la realidad dividiéndola en componentes analizables para el estudio científico, debido a las limitaciones en la observación y el conocimiento [35].

Esta investigación es neopositivista, ya que su finalidad es evaluar el desempeño de un sistema de navegación asistido por visión computacional a partir del análisis cuantitativo de sus módulos funcionales. Se busca medir de forma objetiva las variables técnicas del sistema mediante pruebas en entornos controlados. Este enfoque permite analizar de manera sistemática el desempeño del sistema como un prototipo funcional en condiciones controladas.

3.3 Método

Esta investigación utiliza el método científico [36] para evaluar el desempeño de un sistema de navegación asistido por visión computacional, basado en el algoritmo YOLOv8, técnicas de estereovisión con OpenCV y retroalimentación auditiva mediante Google Text-to-Speech. El método se orienta a la validación técnica del sistema en etapa de prototipo, mediante pruebas controladas que permiten analizar su comportamiento en términos de precisión, velocidad y costo computacional, con el fin de caracterizar el desempeño del sistema de manera integral.

3.4 Tipo de investigación

La investigación es de tipo descriptiva, ya que su finalidad es caracterizar el desempeño de un sistema de navegación asistido por visión computacional en entornos interiores controlados.

Para ello, se establecen parámetros clave del sistema, considerando aspectos como la detección y clasificación de objetos, la estimación de distancias, la asignación espacial y la velocidad de procesamiento, los cuales influyen directamente en el comportamiento general del sistema.

El objetivo es analizar el desempeño del sistema en términos de precisión, velocidad y costo computacional, determinando su funcionamiento como prototipo en condiciones controladas. Esto permite identificar fortalezas, limitaciones y posibles mejoras del sistema para futuras implementaciones en contextos reales.

3.5 Diseño de investigación

El diseño de esta investigación corresponde a un experimento con grupos de comparación, orientado a evaluar la efectividad de un sistema de navegación asistido por visión

computacional en entornos interiores controlados. La evaluación se estructura en dos niveles complementarios: una validación técnica del sistema por módulos, y una comparación funcional del desempeño de navegación con y sin el sistema.

RG1 **X1** **O1**
RG2 **X2** **O2**

Donde:

RG1: Usuario con simulación de discapacidad visual mediante oclusión ocular, realizando recorridos en entorno controlado sin el sistema de navegación asistida.

RG2: Usuario con simulación de discapacidad visual mediante oclusión ocular, realizando recorridos en entorno controlado con el sistema de navegación asistida activo.

X1: Ausencia del sistema de navegación asistida (condición base).

X2: Sistema de navegación asistida integrado, compuesto por YOLOv8, estereovisión con OpenCV y retroalimentación auditiva mediante Google Text-to-Speech.

O1: Número de colisiones, tiempo de recorrido y capacidad de completar la ruta sin asistencia tecnológica.

O2 Número de colisiones, tiempo de recorrido y capacidad de completar la ruta con asistencia del sistema, en términos de precisión, velocidad y costo computacional.

3.6 Universo

El universo de estudio comprende un sistema de navegación asistido por visión computacional que utiliza algoritmos avanzados de visión computacional. Este universo es finito y accesible, ya que se pueden recopilar y analizar datos específicos sobre el desempeño del sistema en diferentes escenarios.

3.7 Muestra

La muestra de estudio coincide con el universo, ya que el análisis se realizará sobre el sistema de navegación asistido por visión computacional en su totalidad. Este sistema, que utiliza el algoritmo YOLOv8 y estereovisión con OpenCV para la detección de objetos y la estimación de profundidad, respectivamente, fue evaluado en diversos escenarios interiores de la Universidad CESMAG. Debido a que el universo es finito y completamente accesible, se recopilan y analizan todos los datos generados por el sistema en distintos entornos, lo que

permitirá obtener conclusiones sobre su desempeño en términos de precisión, velocidad de procesamiento y costo computacional.

3.8 Técnicas de recolección de información

Observación directa: Para la adquisición de datos, se emplean equipos como el módulo de cámara 3D GXIVISION 4MP, que permiten capturar imágenes y videos del entorno. Además, se utiliza software de programación como Python para la implementación del sistema de visión por computador, el cual integra algoritmos como YOLOv8 y técnicas de estereovisión con OpenCV para el procesamiento de la información visual.

La observación se lleva a cabo en escenarios controlados para evaluar el desempeño del sistema de navegación asistido por visión computacional mediante métricas objetivas asociadas a la percepción del entorno, la estimación de distancia, la asignación espacial, la velocidad de procesamiento y el costo computacional.

3.8.1 Validez de la técnica

La técnica de recolección de información es válida, ya que se utilizan equipos especializados, como cámaras, que permitirán registrar métricas objetivas relacionadas con la detección de objetos, la clasificación, la estimación de distancia, la asignación espacial, la velocidad de procesamiento y el costo computacional. Estas métricas reflejan directamente el desempeño del sistema en condiciones controladas.

3.8.2 Confiabilidad técnica

La técnica es confiable, dado que se emplean cámaras para capturar datos en escenarios consistentes y controlados. Esto permite comparar los resultados en diferentes situaciones. Además, se implementan pruebas repetidas con el uso de los mismos equipos, asegurando que los datos obtenidos sean consistentes y representativos.

Estos cambios refuerzan la idea de que los equipos, como las cámaras, son las herramientas clave para la adquisición de datos y aseguran tanto la validez como la confiabilidad del proceso.

3.9 Instrumentos de recolección de información

Los instrumentos de recolección de información están diseñados para capturar datos relevantes que permitirán evaluar la efectividad del sistema de navegación asistido por visión computacional. Estos instrumentos incluyen software Python para la implementación y ejecución del algoritmo de visión computacional YOLOv8 y estereovisión con OpenCV. Sistemas de captura de imágenes y videos en tiempo real como las cámaras con una resolución igual o mayor a 4MP como el módulo de cámara 3D GXIVISION 4MP. Además, se implementan métricas de análisis de datos para medir la precisión, velocidad de detección y costo computacional para evaluar la efectividad del sistema en condiciones controladas.

3.10 Establecimiento de requerimientos para el sistema

Con el fin de cumplir los objetivos planteados para el desarrollo del sistema de navegación asistido por visión computacional, se definieron los requerimientos funcionales y no funcionales que guiaron el diseño, implementación y validación del prototipo. Estos requerimientos se establecieron considerando el uso de algoritmos avanzados de detección (YOLOv8), técnicas de estereovisión mediante OpenCV, mecanismos de retroalimentación auditiva basada en Google Text-to-Speech (gTTS), y la necesidad de contar con una interfaz gráfica para monitoreo en tiempo real.

- El sistema debe detectar objetos del entorno (sillas, mesas, puertas, paredes, señales, botes de basura, entre otros) utilizando el modelo YOLOv8 entrenado con el dataset propio.
- El sistema debe calcular la distancia de los objetos detectados usando un módulo de doble lente calibrado y los algoritmos de disparidad de OpenCV.
- El sistema debe emitir retroalimentación auditiva basada en gTTS cuando se identifiquen objetos a una distancia crítica o cuando se detecten obstáculos no reconocidos mediante la capa de *occupancy grid*.
- El sistema debe permitir la observación del comportamiento del modelo y los eventos generados para facilitar la evaluación por parte de los investigadores.
- El sistema debe contar con una interfaz gráfica en el computador que muestre detecciones, distancias, instrucciones y FPS del sistema en tiempo real.

3.11 Diseño del sistema

El prototipo desarrollado para la navegación asistida mediante visión computacional está compuesto por un sistema de cámaras estereoscópicas, algoritmos de reconocimiento y estimación de distancias, y un módulo de retroalimentación auditiva. El diseño integra componentes de hardware y software que permiten detectar objetos relevantes del entorno, estimar su proximidad y comunicar esta información al usuario mediante instrucciones de voz en tiempo real.

En la implementación actual, el módulo de cámara estéreo se monta sobre una correa ajustada a la cintura del usuario, orientada hacia la dirección de avance. El procesamiento se lleva a cabo en un computador portátil, al cual la cámara se conecta vía USB. Esta configuración permitió ejecutar los modelos de visión artificial en tiempo real y realizar las pruebas de navegación en condiciones controladas, manteniendo estabilidad en el procesamiento y en la captura de datos.

3.11.1 Plataforma de hardware utilizada

Para las pruebas experimentales, el sistema se ejecutó en un computador portátil con las siguientes especificaciones:

- Procesador: AMD Ryzen 7 7735HS (8 cores, 16 threads)
- Memoria RAM: 16 GB
- GPU dedicada: NVIDIA GeForce RTX 4060 (8 GB VRAM)
- GPU integrada: AMD Radeon Graphics
- Sistema operativo: Windows 11
- Conexión del módulo de cámaras: USB 2.0

Estas características permiten evaluar el costo computacional, el rendimiento del modelo YOLOv8 y la eficiencia del proceso de estereovisión bajo una plataforma moderna, aunque no optimizada específicamente para dispositivos portátiles finales.

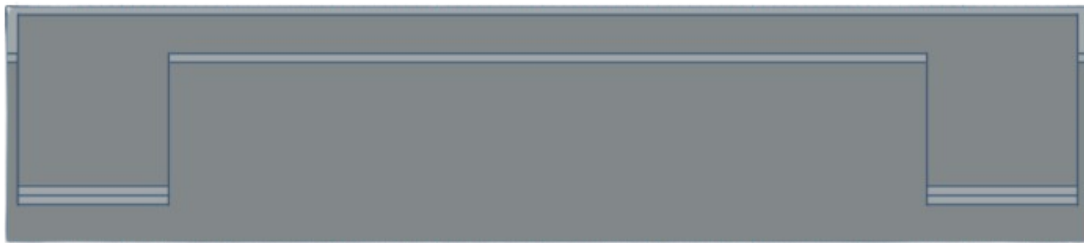
3.11.2 Diseño 3D del soporte para el módulo de cámaras estéreo

El soporte estructural del módulo de cámaras se diseñó en Tinkercad, buscando una base liviana y estable. El modelo 3D presenta una estructura sencilla y funcional, adaptada a las

dimensiones del módulo y permite insertarlo fácilmente por deslizamiento. Se pensó para impresión 3D en material PLA, por su resistencia y bajo peso. Se añadió una tira elástica ajustable que permite sujetar el sistema al abdomen del usuario, asegurando comodidad y estabilidad durante el desplazamiento. Su forma compacta evita interferir con los movimientos de la persona. El modelo se exportó en formato .STL para su impresión o modificación. Este soporte garantiza una captura de imágenes estable.

Figura 10.

Diseño 3D del soporte



3.11.3 Conexión del módulo de doble lente con el computador

La integración del módulo de estereovisión con el computador se realizó mediante una conexión directa por cable USB, que permite la transmisión simultánea de video desde ambos lentes. Esta conexión facilita la comunicación entre el módulo y el sistema de procesamiento.

El uso del puerto USB simplifica la instalación, evita configuraciones adicionales y garantiza una transferencia de datos estable y de baja latencia. Gracias a esta conexión única, el sistema puede capturar, procesar y analizar imágenes en tiempo real, manteniendo una comunicación continua.

3.11.4 Configuración del módulo de visión

El módulo de visión está conformado por una cámara estéreo USB de doble lente (4 MP, 1080p). Este módulo permite capturar imágenes simultáneas desde dos perspectivas ligeramente separadas, requisito indispensable para la estereovisión.

Para la calibración se utilizó el método de patrón de ajedrez en OpenCV, lo que permitió calcular los parámetros intrínsecos (matriz de cámara, coeficientes de distorsión) y extrínsecos (rotación y traslación entre cámaras). Con estos parámetros se corrigió la distorsión radial y tangencial, y se obtuvo la matriz de disparidad.

La disparidad d se calculó como:

$$d = x_L - x_R \quad (3)$$

donde:

x_L es la coordenada horizontal del punto en la cámara izquierda.

x_R es la coordenada horizontal del punto en la cámara derecha.

La distancia Z al objeto se obtuvo aplicando la ecuación:

$$Z = \frac{f \cdot B}{d} \quad (4)$$

donde:

f es la distancia focal de la cámara.

B es la distancia entre los centros ópticos de las dos cámaras (baseline).

d es la disparidad previamente calculada.

Figura 11

Patrón captado por ambos lentes (izquierdo y derecho) para calibración de estereovisión



3.11.5 Diseño del dataset

Para el sistema de detección se construyó un conjunto de datos propio, integrado por 29 clases que representan objetos críticos para la orientación y seguridad en el entorno universitario.

A cada categoría se le asignó un identificador numérico (0-28), facilitando la trazabilidad durante las etapas de anotación y entrenamiento. La relación completa de etiquetas y su nomenclatura técnica se detalla en la Tabla 2.

La construcción del dataset partió de la captura de 100 imágenes originales por clase. Con el fin de robustecer el modelo frente a condiciones ambientales variables, se aplicaron técnicas de aumentación de datos (rotaciones, ajustes de iluminación, escalado y recortes), incrementando el volumen a 1000 imágenes por clase, para un total de 29,000 muestras. El etiquetado se realizó mediante la herramienta LabelImg, tal como se muestra en la figura 12, empleando cajas delimitadoras (*bounding boxes*) bajo el estándar de YOLOv8. Esta estructura de datos fue diseñada para permitir una implementación flexible, soportando tanto el entrenamiento integral como esquemas de ejecución modular para clases estructurales y objetos de interés.

Tabla 2

Clases del dataset utilizadas para el entrenamiento del modelo YOLOV8n

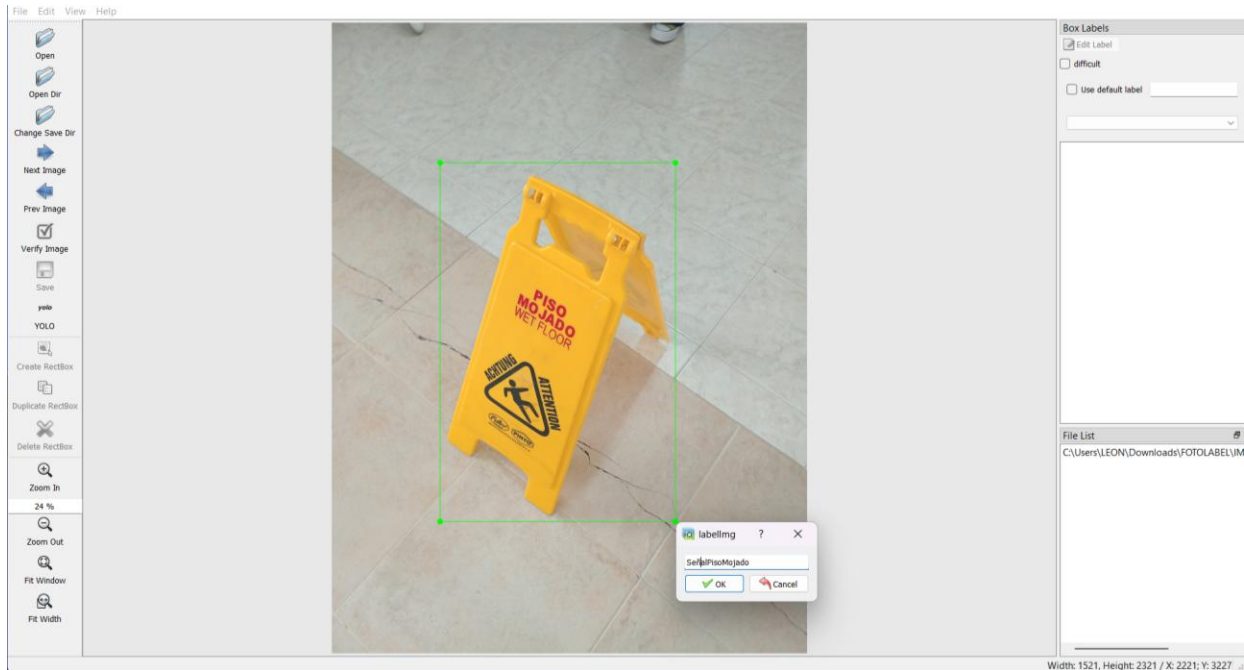
ID	Nombre de clase en dataset	Nombre genérico
0	Generador	Generador
1	Osciloscopio	Osciloscopio
2	FuenteDual	Fuente de poder dual
3	MesaConSillas	Conjunto de mesa y sillas
4	Botedebasura	Contenedor de basura
5	Extintor	Extintor de incendios
6	Botedereciclaje	Contenedor de reciclaje
7	ProhibidoPV	Prohibido el paso vehicular
8	PinPongMesa	Mesa de ping-pong
9	BateriaBoteReciclaje	Contenedor de baterías
10	AtwoodMachine	Máquina de Atwood
11	SeñalPisoMojado	Señal de piso mojado
12	CamillaEmergencia	Camilla de emergencia
13	Gradas	Gradas
14	CNCPCB	Máquina CNC para PCB

15	BotellasPlastico	Contenedor de botellas de plástico
16	PuestoBicicletas	Puesto de bicicletas
17	Escaleras	Escaleras
18	Elevador	Elevador
19	Barandales	Barandales
20	EntradaRFID	Torniquete de acceso
21	Silla	Silla
22	Puerta	Puerta
23	Mueble	Escritorio/Banca/Cajonera
24	MáquinaDispensadora	Máquina Expendedora
25	Computador	Computador
26	Casillero	Casillero
27	Mesa	Mesa
28	Wall	Pared/Muro

Es importante señalar que la categoría 'Mueble' se diseñó como una clase compuesta. En ella se agruparon objetos con características morfológicas similares, tales como escritorios, cajoneras y bancas. Esta agrupación estratégica permite que el sistema mantenga una alta capacidad de generalización ante obstáculos de dimensiones similares, garantizando una respuesta consistente para el usuario.

Figura 12

Ejemplo de objeto siendo etiquetado con LabelImg



3.11.6 Entrenamiento del modelo YOLOv8

Para el módulo de detección se implementó el modelo YOLOv8n, una versión optimizada que ofrece un equilibrio adecuado entre velocidad de inferencia y precisión para aplicaciones en tiempo real. El entrenamiento se basó en el dataset previamente descrito, empleando una distribución de 80 % para entrenamiento y 20 % para validación. El proceso completo de optimización se ejecutó durante 32.24 horas, obteniéndose los pesos finales integrados en el sistema de navegación.

Con el objetivo de maximizar la fiabilidad del sistema, se adoptó una arquitectura modular de detección en paralelo. Bajo este enfoque, las primeras 28 clases fueron entrenadas de forma conjunta, mientras que la clase estructural 'Wall' (identificada como ID 28 en el dataset maestro) fue procesada mediante una instancia independiente. En dicha instancia, la clase asume el ID 0 por configuración técnica del modelo unicategoría, manteniendo la integridad de las etiquetas originales en la fase de inferencia. Esta especialización permitió alcanzar una sensibilidad máxima en la detección de límites físicos, asegurando que la identificación de obstáculos críticos no se vea afectada por la varianza estadística de objetos de menor tamaño.

3.11.7 Diseño de la retroalimentación auditiva

Para la síntesis de voz del sistema se utilizó Google Text-to-Speech (gTTS) como motor principal. Esta herramienta permitió generar instrucciones claras y dinámicas, con un procesamiento ligero y un manejo estable de colas de mensajes, evitando que las alertas se superpusieran o se reprodujeran de forma simultánea.

Su principal ventaja es su eficiencia: requiere muy pocos recursos y funciona con buena latencia incluso en tiempo real. La única limitación relevante es la necesidad de contar con conexión a internet para generar las locuciones, pero debido a su rendimiento y estabilidad se consideró la mejor alternativa para el sistema final.

También se hace uso de audífonos inalámbricos de conducción aérea, con el propósito de evitar saturar el oído del usuario y que tenga la capacidad de poder escuchar su entorno y el sistema al mismo tiempo.

Figura 13

Audífonos inalámbricos de conducción aérea.



3.11.8 Diseño de las instrucciones

Las instrucciones de voz están estructuradas para ofrecer al usuario información clara y útil en situaciones de movilidad. Cada mensaje incluye:

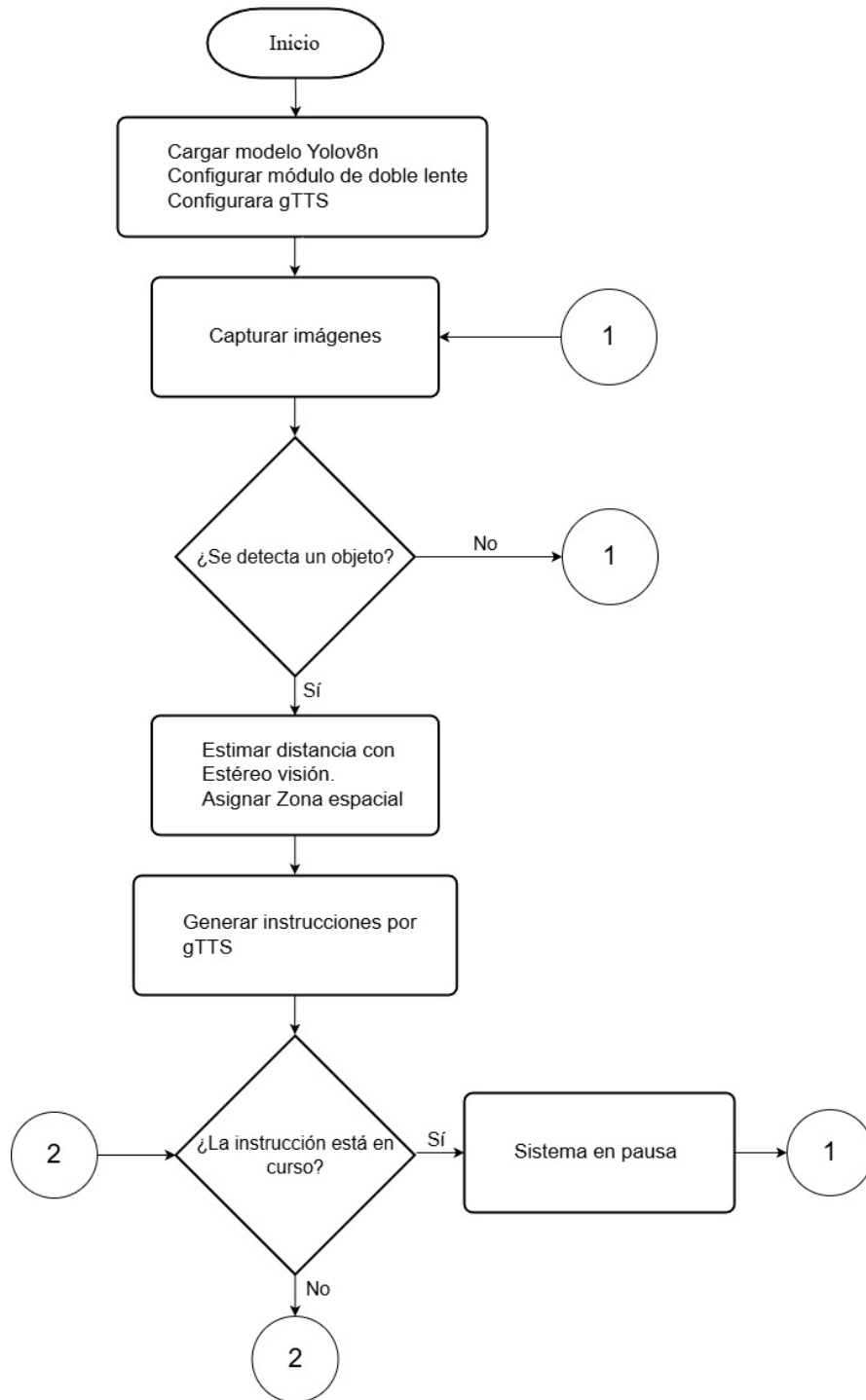
- **Nombre del objeto detectado y clasificado** mediante YOLOv8n.

- **Distancia estimada en metros**, obtenida a partir de la estereovisión.
- **Dirección expresada en formato de zonas horarias** (por ejemplo: “a las once”, “a las doce”, “a la una”), lo cual permite al usuario ubicar objetos sin necesidad de referencia visual.

Adicionalmente, se implementó un mecanismo de control de emisión para evitar la repetición continua de instrucciones en frames consecutivos. Dado que el modelo YOLOv8n realiza detecciones de manera constante, un mismo objeto podría generar múltiples alertas sucesivas. Para prevenir solapamientos o interrupciones en la reproducción del mensaje, el sistema bloquea temporalmente la detección mientras se encuentra activa la síntesis de voz. Una vez finalizada la instrucción, el proceso de inferencia se reanuda inmediatamente. Este control garantiza la entrega completa de cada mensaje al usuario.

Figura 14.

Diagrama de flujo de generación de mensajes de voz.



3.11.9 Configuración inicial del módulo del sistema

La integración inicial del sistema consistió en unificar los módulos principales, detección de objetos con YOLOv8, estimación de profundidad mediante técnicas de estereovisión con OpenCV y retroalimentación auditiva a través de Google Text-to-Speech (gTTS).

En una primera etapa, cada módulo fue evaluado de forma independiente para comprobar su funcionamiento y rendimiento. Posteriormente, se desarrolló un programa central en Python que recibe las imágenes del módulo de cámaras, procesa la detección de objetos, calcula su distancia en tiempo real y finalmente genera una instrucción auditiva para el usuario.

Durante este proceso se verificó la correcta sincronización entre la detección, la estereovisión y la generación de audio, asegurando un flujo continuo de información sin interrupciones. Para evitar saturación de mensajes o solapamientos en la reproducción, se implementó un mecanismo de control mediante el cual el sistema entra en pausa mientras gTTS reproduce una instrucción, reanudando automáticamente la detección al finalizar el mensaje. Este comportamiento garantiza que el usuario reciba instrucciones claras, secuenciales y sin confusión auditiva.

Estas pruebas iniciales permitieron comprobar la comunicación adecuada entre todos los componentes y sentaron las bases para el funcionamiento del sistema de navegación.

3.12 Implementación del sistema

3.12.1 Integración del hardware

La integración del hardware del sistema se realizó estableciendo una arquitectura centralizada en el computador portátil, el cual actúa como nodo de procesamiento principal. El módulo de cámaras estéreo USB de doble lente se conecta directamente al computador mediante una interfaz USB, permitiendo la adquisición sincronizada de imágenes para los módulos de visión y estereovisión.

Por su parte, la salida de retroalimentación auditiva se transmite desde el computador hacia los audífonos de conducción aérea, los cuales reciben las instrucciones generadas por el sistema de síntesis de voz. Esta integración permite mantener un flujo continuo de información desde la captura visual hasta la emisión de mensajes auditivos, garantizando la operación coordinada del sistema.

La Figura 15 presenta el esquema general de interconexión entre los componentes de hardware y los módulos de software, mientras que la Figura 16 muestra el montaje físico del prototipo en su configuración portátil.

Figura 15

Diagrama de flujo de conexión entre hardware y software.

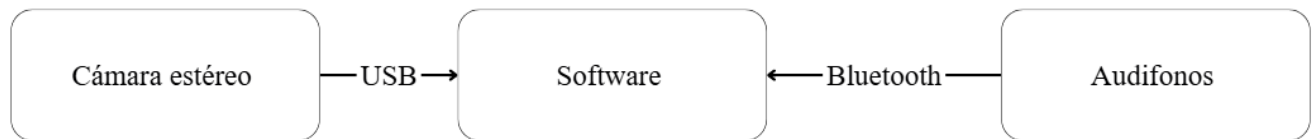


Figura 16

Configuración física del prototipo en su diseño portátil.



3.12.2 Integración de software

Los módulos de visión, estereovisión y generación de instrucciones auditivas fueron integrados en un sistema desarrollado en Python. Para su implementación se utilizaron librerías especializadas como OpenCV para el procesamiento de imágenes, PyTorch como base del

modelo de detección YOLOv8 a través del framework Ultralytics, y Google Text-to-Speech (gTTS) para la generación de mensajes auditivos.

La comunicación entre módulos permite realizar la detección de objetos, la estimación de distancias y la generación de mensajes auditivos en un mismo flujo de trabajo. Durante la ejecución, el sistema captura imágenes desde el módulo de cámaras estéreo, procesa la información para detectar objetos y estimar su distancia, y posteriormente genera instrucciones auditivas dirigidas al usuario.

La reproducción de instrucciones por voz pausa temporalmente el procesamiento visual con el fin de evitar la superposición de mensajes; una vez finalizada cada instrucción, el sistema reanuda automáticamente la detección y el cálculo de distancias.

3.12.3 Interfaz de monitoreo

La interfaz de monitoreo del sistema fue desarrollada utilizando PySide6, permitiendo una visualización más moderna y organizada que la ofrecida por la ventana nativa de OpenCV.

En ella se muestra en tiempo real el flujo de video proveniente de las cámaras estéreo, donde se superponen las detecciones generadas por YOLOv8 y las distancias estimadas mediante estereovisión. La interfaz integra además un panel informativo donde se registran los mensajes auditivos emitidos por el sistema y FPS. La actualización de todos estos elementos se gestiona mediante temporizadores (QTimer), lo que permite mantener una visualización continua sin bloquear el hilo principal de ejecución tal como se muestra en la Figura 22.

3.12.4 Configuración del prototipo y modo de operación

El sistema completo estuvo conformado por un módulo de cámaras estéreo USB de doble lente, un computador portátil encargado del procesamiento en tiempo real y audífonos de conducción aérea utilizados para la retroalimentación auditiva mediante el sistema de síntesis de voz Google Text-to-Speech (gTTS). El prototipo fue diseñado para operar de forma portátil, incluyendo una base impresa en 3D para el montaje del módulo estéreo y la posibilidad de funcionamiento con la tapa del computador cerrada.

No obstante, durante las pruebas experimentales el computador fue operado de manera supervisada por los investigadores, con el fin de monitorear continuamente la interfaz gráfica del sistema y registrar los resultados obtenidos. Los desplazamientos se realizaron de forma

controlada, simulando el recorrido de una persona con discapacidad visual, sin involucrar a terceros.

3.13 Protocolo experimental y criterios de evaluación

Las pruebas experimentales se realizaron en entornos controlados dentro de las instalaciones de la Universidad CESMAG, con el fin de evaluar el desempeño del sistema de navegación asistido. Para ello, se definieron tres escenarios distintos, seleccionados por presentar configuraciones espaciales y distribuciones de obstáculos diferentes. En todos los escenarios se estableció una ruta general de desplazamiento (Punto A → Punto B), en la cual se dispusieron instancias reales de los objetos considerados en el dataset. Si bien el punto de inicio y final se mantuvo constante, el trayecto específico pudo variar en función de las decisiones tomadas por el usuario durante el desplazamiento, lo que implicó que no todos los objetos fueran enfrentados en cada recorrido. En total, se realizaron 20 recorridos por escenario, durante los cuales el sistema ejecutó de manera simultánea la detección y clasificación de objetos mediante YOLOv8, la estimación de distancias a través de estereovisión y la retroalimentación auditiva utilizando Google Text-to-Speech.

Durante las pruebas, el desplazamiento se realizó a velocidad reducida y de manera progresiva, permitiendo al sistema procesar la información y emitir instrucciones antes de que el usuario avanzara significativamente. Esta condición se estableció con el fin de evaluar el desempeño del prototipo en un entorno controlado, priorizando la seguridad y la correcta validación de los módulos integrados.

3.13.1 Criterios de evaluación por módulo

Con el fin de evaluar de manera estructurada el desempeño del sistema propuesto, se definieron criterios de evaluación independientes para cada uno de sus módulos principales, así como un criterio integrador a nivel del sistema completo.

- **Detección (criterio A):** Evalúa si un objeto presente en la escena fue detectado por el modelo, independientemente de si la clase asignada fue correcta. Para este criterio, el módulo de detección opera con un umbral de confianza mínimo, priorizando la sensibilidad (Recall) con el fin de garantizar la detección de todos los obstáculos potenciales.

- **Clasificación (criterio B):** Evalúa si el objeto detectado fue identificado con su etiqueta correcta en al menos un frame durante la interacción. Se considera un acierto de clasificación cuando la clase correcta es asignada al objeto en cualquier instante del proceso de observación.
- **Estimación de distancia (criterio C):** Evalúa la precisión de la distancia estimada mediante el error absoluto entre la distancia real y la estimada. Se establece un umbral de tolerancia de ± 10 cm para considerar una estimación como correcta. Las distancias reales fueron medidas utilizando un flexómetro metálico con una precisión de ± 2 mm, desde la lente de la cámara hasta el punto más cercano del objeto.
- **Asignación de zona espacial correcta (Criterio D):** Evalúa la correcta asignación de la zona espacial del objeto dentro del campo de visión horizontal de la cámara. El espacio observable se divide en siete zonas discretas (9, 10, 11, 12, 1, 2 y 3), que corresponden a posiciones espaciales percibidas por el usuario. La zona correcta se determina de forma geométrica a partir de la posición horizontal del centroide del bounding box, normalizada respecto al ancho de la imagen, y se compara con la zona reportada por el sistema. Una asignación se considera correcta cuando ambas coinciden. Este criterio se evalúa exclusivamente sobre objetos detectados, dado que la asignación de zona depende directamente de la existencia de una detección previa.
- **Éxito del sistema completo (end-to-end):** La evaluación end-to-end del sistema se realizó bajo dos condiciones complementarias. En la primera, correspondiente al criterio de seguridad, un recorrido se considera exitoso únicamente cuando el usuario completa el trayecto sin colisiones con los obstáculos presentes en el entorno; la ocurrencia de una o más colisiones implica la clasificación del recorrido como no exitoso, incluso si se alcanza el punto final, permitiendo así evaluar la robustez del sistema bajo condiciones de seguridad ideal. En la segunda condición, de carácter comparativo, se permite la ocurrencia de colisiones durante el recorrido sin que estas interrumpan la prueba, con el fin de analizar el comportamiento del usuario en escenarios con y sin asistencia del sistema; en este caso, las colisiones no determinan el éxito del recorrido, sino que se registran como una variable de desempeño asociada a la seguridad de navegación, permitiendo evaluar el impacto del sistema en métricas como tiempo de recorrido, número de colisiones y finalización del trayecto.

Adicionalmente, se evaluaron métricas complementarias de desempeño:

- **Velocidad de instrucciones de voz:** Tiempo transcurrido entre la detección del objeto y la emisión de la instrucción auditiva correspondiente al usuario.
- **Velocidad de procesamiento:** Número de frames por segundo (FPS) procesados durante la ejecución del sistema.

Sobre estos criterios se aplicaron las métricas cuantitativas definidas en la Sección

[2.3.2]:

- Para detección: Recall.
- Para clasificación: Accuracy.
- Para estimación de distancia: Accuracy basada en umbral.
- Para asignación de zona espacial: Accuracy.
- Y para el sistema completo (end-to-end): Precision, Recall y F1-Score.

4 Resultados

4.1 Prototipo del módulo de cámaras instalado en el sistema

Como resultado del proceso de diseño e impresión del soporte, se procedió al montaje del módulo de cámaras estéreo junto con la correa de sujeción que permite fijar el sistema al usuario durante la navegación. Esta configuración constituye el prototipo físico utilizado en las pruebas experimentales del sistema de asistencia basado en visión computacional.

Figura 17

Módulo de doble lente para estereovisión junto a la correa y el soporte



4.2 Resultados de detección del modelo

Una vez finalizado el proceso de entrenamiento del modelo YOLOv8, se realizaron pruebas de validación utilizando el conjunto de datos preparado previamente. Durante esta etapa, el algoritmo generó visualizaciones automáticas que permiten observar el comportamiento del modelo al identificar los objetos presentes en las imágenes.

La Figura 18 muestra mosaicos generados durante la fase de validación, en los cuales se pueden observar distintas muestras del dataset junto con las cajas delimitadoras (bounding boxes) y las etiquetas asignadas a cada objeto detectado. Estas visualizaciones permiten verificar que el modelo reconoce correctamente las diferentes clases incluidas en el conjunto de entrenamiento.

e)



f)



Nota. La imagen presenta mosaicos de datos generados automáticamente por el algoritmo durante la fase de validación. Se observan las etiquetas e identificadores (ID) asignados a las 29 clases del dataset, incluyendo cajas delimitadoras (bounding boxes) que permiten visualizar la localización y clasificación de los objetos y elementos estructurales.

4.3 Resultados del entrenamiento del modelo

Durante la fase de validación, el modelo generó métricas que confirman su estabilidad global. Las curvas Recall–Confidence, Precision–Confidence, F1–Confidence y Precision–Recall (Figura 19) muestran valores altos y consistentes, lo que indica una adecuada generalización y ausencia de sobreajuste. Esta capacidad de detección se refleja visualmente en los mosaicos de validación presentados anteriormente, donde el modelo logra identificar y delimitar los objetos con precisión basándose en los identificadores (ID) asignados.

Finalmente, los resultados cuantitativos se resumen en la Tabla 3. Estos valores de Precisión, Recall y F1-score proporcionan la base estadística que valida la efectividad del sistema para su despliegue en entornos de navegación asistida.

Tabla 3*Métricas de desempeño por clase para el sistema de detección modular YOLOv8n*

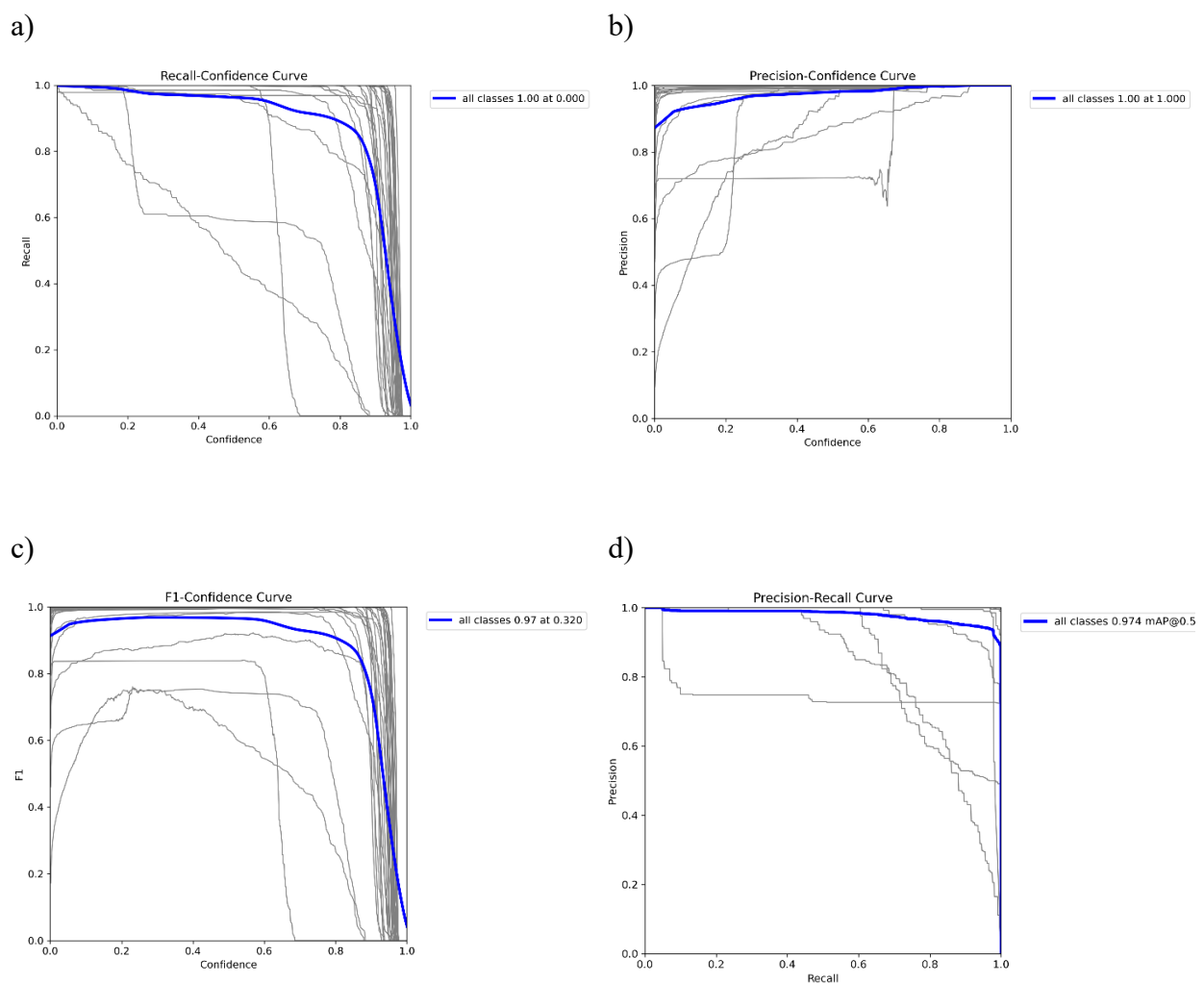
ID	TP	FP	FN	Precision (%)	Recall (%)	F1-Score (%)
0	198	0	0	100.00	100.00	100.00
1	191	1	0	99.50	100.00	99.74
2	208	0	0	100.00	100.00	100.00
3	202	0	0	100.00	100.00	100.00
4	203	0	0	100.00	100.00	100.00
5	201	0	0	100.00	100.00	100.00
6	197	0	0	100.00	100.00	100.00
7	200	0	0	100.00	100.00	100.00
8	200	0	0	100.00	100.00	100.00
9	200	0	0	100.00	100.00	100.00
10	196	51	4	79.35	98.00	87.69
11	200	0	0	100.00	100.00	100.00
12	200	0	0	100.00	100.00	100.00
13	200	77	0	72.20	100.00	83.86
14	200	0	0	100.00	100.00	100.00
15	200	0	0	100.00	100.00	100.00
16	200	0	0	100.00	100.00	100.00
17	122	3	78	97.60	61.00	75.08
18	200	0	0	100.00	100.00	100.00
19	150	31	50	82.87	75.00	78.74
20	200	2	0	99.00	100.00	99.50
21	225	8	5	96.57	97.83	97.19
22	254	4	0	98.45	100.00	99.22
23	241	1	0	99.59	100.00	99.79
24	200	0	0	100.00	100.00	100.00
25	200	0	0	100.00	100.00	100.00
26	93	0	0	100.00	100.00	100.00

27	200	2	0	99.00	100.00	99.50
28	120	10	0	92.30	100.00	96.00

Nota. Los valores de rendimiento se presentan en formato porcentual para mantener la coherencia con las pruebas de efectividad del sistema. El desglose detallado de las interacciones para la instancia multiobjeto (ID 0-27) se presenta en la matriz de confusión normalizada (ver Anexo A); en ella se observa una precisión predominante en objetos de mobiliario, con variaciones esperadas en elementos de alta complejidad geométrica como barandales y escaleras. Los resultados de la instancia estructural se documentan en el Anexo B.

Figura 19

Curvas de rendimiento del modelo YOLOv8n durante la fase de validación



Nota. Evaluación del desempeño para las categorías de objetos (ID 0-27) mediante: (a) curva Recall-Confidence; (b) curva Precision-Confidence; (c) curva F1-Confidence; (d) curva Precision-Recall. Los resultados de la instancia estructural (clase Wall) se omiten visualmente por presentar una convergencia lineal hacia la unidad (1.00), derivada de su naturaleza unicategoría y alta especialización, el análisis detallado de esta clase puede consultarse en los Anexos B y C.

4.4 Resultados experimentales del sistema

En esta sección se presentan los resultados obtenidos durante las pruebas experimentales realizadas en los tres escenarios definidos dentro de las instalaciones de la Universidad CESMAG. Las evaluaciones se desarrollaron siguiendo el protocolo experimental descrito en la metodología, donde el sistema ejecutó de manera simultánea los módulos de detección de objetos mediante YOLOv8, estimación de distancia basada en estereovisión y generación de instrucciones auditivas para el usuario. Los resultados se analizan considerando los criterios de evaluación establecidos para cada módulo del sistema, así como el desempeño global del sistema durante los recorridos realizados entre el punto A y el punto B.

4.4.1 Escenario 1: Salón H501

Las primeras pruebas se realizaron en el salón H501 (edificio Holanda, 5.º piso). En este escenario se dispusieron 12 instancias reales de objetos pertenecientes al dataset, incluyendo un mueble, sillas, paredes, una puerta, una señal de piso mojado y botes de basura, distribuidos a lo largo del recorrido definido entre el punto A y el punto B. Durante las evaluaciones se presentaron variaciones de iluminación natural, así como oclusiones parciales derivadas de la disposición de los objetos en el entorno.

En este escenario se realizaron 20 pruebas experimentales, de las cuales en 4 recorridos el sistema no logró completar el trayecto sin colisión, fallos atribuibles principalmente a errores en la estimación de profundidad por estereovisión en determinados objetos y, en menor medida, a detecciones incorrectas por parte del modelo YOLOv8.

En la Figura 21 se presenta el esquema del escenario, donde se indican la ruta de desplazamiento, la ubicación de los objetos y las dimensiones del espacio.

Figura 20

Vista panorámica del entorno experimental: Salón H501 (Escenario 1)



Figura 21

Esquema del escenario 1 (Salón H501) con la ruta de desplazamiento y disposición de objetos

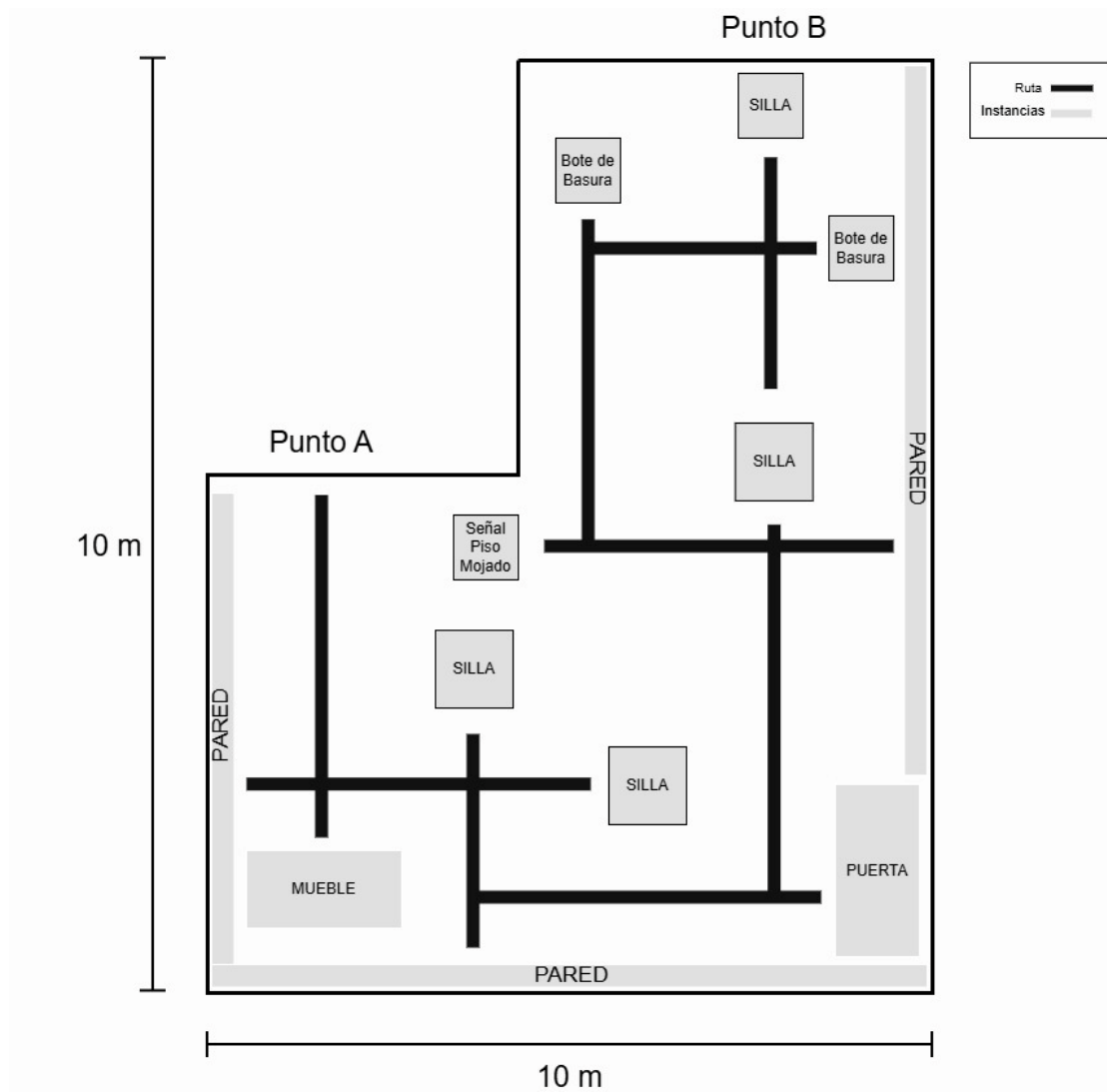


Tabla 4

Detección de objetos (Criterio A)

Prueba	Objetos enfrentados	Objetos detectados (TP)	No detectados (FN)	Recall (%)
1	12	12	0	100.00
2	11	11	0	100.00
3	11	11	0	100.00

4	12	12	0	100.00
5	3	2	1	66.67
6	12	12	0	100.00
7	11	11	0	100.00
8	12	12	0	100.00
9	5	4	1	80.00
10	10	10	0	100.00
11	10	10	0	100.00
12	10	10	0	100.00
13	10	10	0	100.00
14	12	12	0	100.00
15	12	11	1	91.67
16	12	12	0	100.00
17	11	11	0	100.00
18	9	9	0	100.00
19	10	10	0	100.00
20	11	11	0	100.00
TOTAL	206	203	3	98.54

El sistema de detección basado en YOLOv8 mostró un desempeño robusto en las rutas evaluadas. De un total de 206 objetos presentes en las 20 pruebas realizadas, el modelo detectó correctamente 203 instancias, alcanzando un recall del 98.54%.

Solo se presentaron fallos de detección en cuatro pruebas específicas: una silla no detectada en la prueba 5, una puerta en la prueba 9 y una silla en la prueba 15. Estas fallas representan únicamente el 1.45% del total de objetos (4/206), y se concentraron en condiciones particulares de iluminación desfavorable o posicionamiento del objeto en el límite del campo visual.

Es importante destacar que no se registraron falsos positivos ($FP = 0$) en ninguna de las pruebas, lo que significa que el sistema nunca detectó objetos inexistentes. Esta característica es crucial para la navegación asistida, ya que garantiza que todas las alertas emitidas al usuario corresponden a obstáculos reales, evitando confusión o desconfianza en el sistema.

En 17 de las 20 pruebas (85%), el modelo detectó la totalidad de objetos presentes, evidenciando alta consistencia en escenarios controlados.

La Figura 22 muestra un ejemplo de detección exitosa (TP), donde el sistema generó un bounding box sobre la silla, independientemente de la exactitud en la clasificación. Por el contrario, la Figura 23 ilustra un caso de fallo de detección (FN) que resultó en colisión durante la Prueba 5.

Figura 22

Ejemplo de objeto detectado (TP)

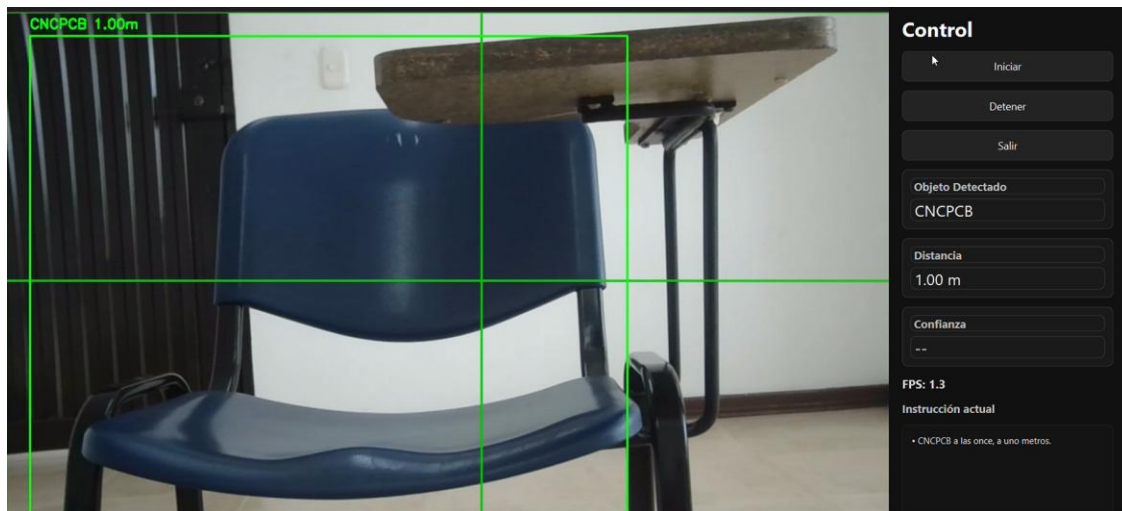


Figura 23

Fallo de detección (FN): silla presente pero no detectada. Colisión en prueba 5

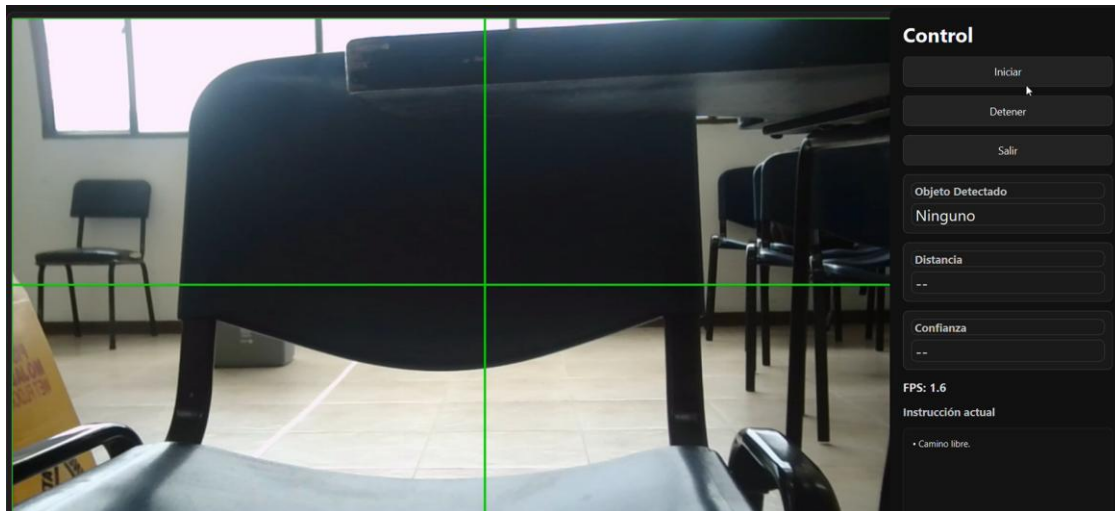


Tabla 5

Desempeño de clasificación de objetos (Criterio B)

Prueba	Objetos detectados (n)	Clasificados correctamente (TP)	Clasificados incorrectamente (FP)	Accuracy (%)
1	12	12	0	100.00
2	11	11	0	100.00
3	11	10	1	91.00
4	12	11	1	91.67
5	2	2	0	100.00
6	12	11	1	91.67
7	11	9	2	81.82
8	12	11	1	91.67
9	4	3	1	75.00
10	10	10	0	100.00
11	10	9	1	90.00
12	10	10	0	100.00

13	10	9	1	90.00
14	12	12	0	100.00
15	11	11	0	100.00
16	12	10	2	83.33
17	11	10	1	91.00
18	9	7	2	77.78
19	10	10	0	100.00
20	11	8	3	72.72
TOTAL	203	186	17	91.62

Para este criterio, se consideró un objeto como clasificado correctamente cuando al menos en un frame durante la interacción el modelo asignó la etiqueta correcta, incluso si en momentos posteriores fluctuó entre clases erróneas. Este enfoque resulta coherente con la aplicación en navegación asistida, donde una clasificación correcta puntual es suficiente para generar instrucciones útiles al usuario.

El modelo demostró un desempeño sólido, de las 203 instancias detectadas en las 20 pruebas realizadas, 186 fueron clasificadas correctamente, alcanzando un accuracy del 91.62%. Por el contrario, los 16 errores de clasificación se concentraron principalmente en condiciones de iluminación desfavorable —contraluces o brillos intensos— que comprometían la percepción visual del objeto e incrementaban la probabilidad de confusión temporal entre clases visualmente similares. Esta limitación se ilustra en la Figura 25, donde el efecto de contraluz genera un silueteado del objeto, ocultando las características visuales necesarias para una clasificación precisa y derivando en una asignación de etiqueta incorrecta.

Figura 24

Validación exitosa de los criterios de clasificación, distancia y zona

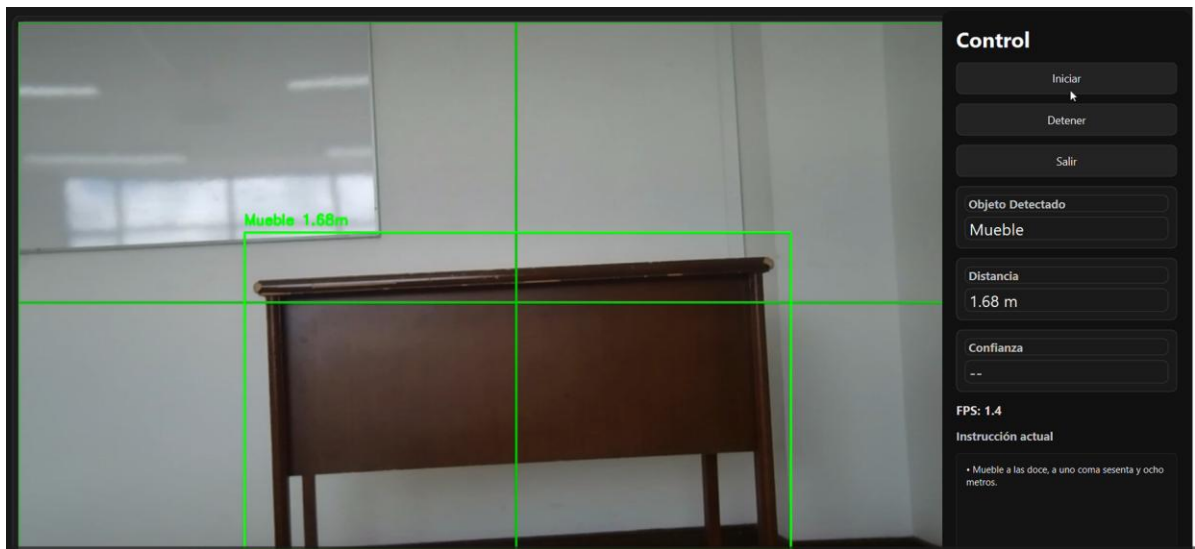


Figura 25

Error de clasificación: bote de basura etiquetado incorrectamente como mueble



Tabla 6*Precisión de la estimación de distancias (Estereovisión) sobre objetos detectados (Criterio C)*

Prueba	Objetos detectados (n)	Objetos estimados correctamente (TP)	Objetos estimados incorrectamente (FP)	Accuracy (%)
1	12	10	2	83.33
2	11	9	2	81.82
3	11	9	2	81.82
4	12	11	1	91.67
5	2	2	0	100.00
6	12	11	1	91.67
7	11	9	2	81.82
8	12	11	1	91.67
9	4	3	1	75.00
10	10	9	1	90.00
11	10	7	3	70.00
12	10	8	2	80.00
13	10	9	1	90.00
14	12	10	2	83.33
15	11	8	3	72.72
16	12	11	1	91.67
17	11	8	3	72.72
18	9	9	0	100.00
19	10	9	1	90.00
20	11	8	3	72.72
TOTAL	203	171	32	84.23

Para este criterio, se consideró una estimación de distancia como correcta cuando el valor calculado por el sistema se encontraba dentro del umbral de ± 10 cm respecto a la distancia real

del objeto. La estereovisión alcanzó un *accuracy* global del 84.23% sobre los objetos efectivamente detectados por el sistema. De las 203 instancias evaluadas a lo largo de las 20 pruebas realizadas, 171 presentaron una estimación correcta; un ejemplo de este desempeño óptimo se observa en la Figura 24, donde el sistema estima con precisión la distancia de un mueble a 1.68 m.

Por otro lado, 32 instancias superaron el umbral definido, clasificándose como errores de precisión métrica. Un caso representativo de esta limitación se ilustra en la Figura 26, donde un objeto situado a una distancia real de 1.69 m fue calculado por el sistema como 1.82 m. Esta desviación de 13 cm, aunque leve, sitúa el registro fuera del margen de tolerancia establecido, impactando directamente en la métrica de efectividad funcional total.

Figura 26

Evaluación del criterio C. La captura ilustra una imprecisión en la estimación de distancia

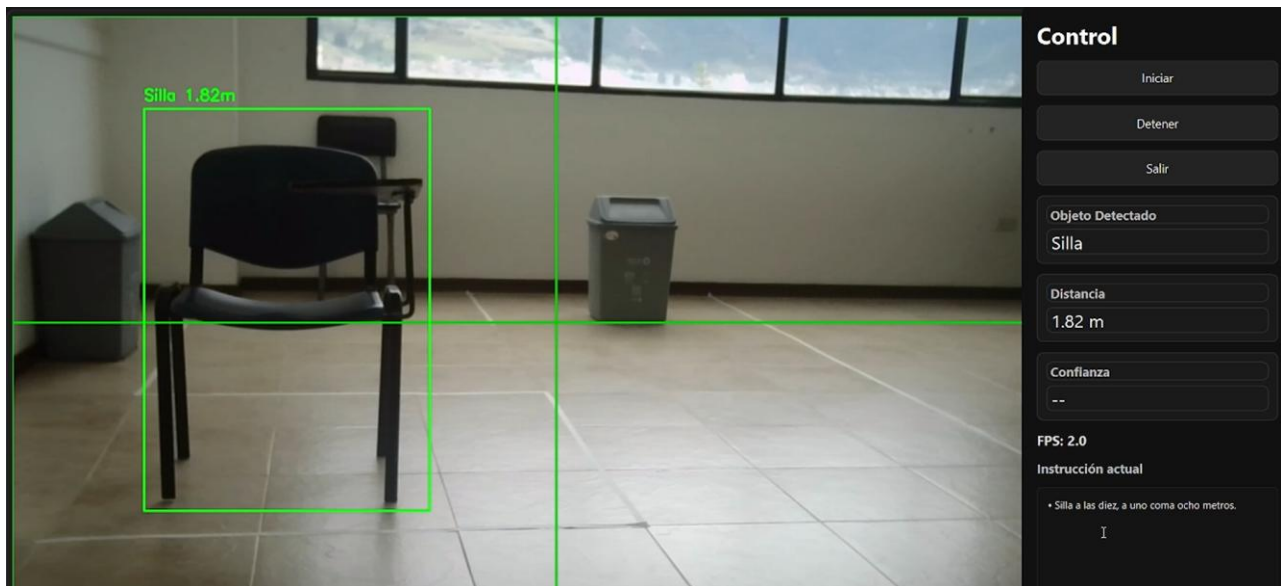


Tabla 7

Resultados de la validación de asignación de zonas espaciales (Criterio D)

Zona	Total	Correctas	Accuracy (%)
9	1	1	100.00
10	4	4	100.00

11	7	7	100.00
12	15	15	100.00
1	7	7	100.00
2	4	4	100.00
3	2	2	100.00
Total	40	40	100.00

Para el Escenario 1, se validó la asignación de zona mediante análisis geométrico retrospectivo de una muestra aleatoria de 40 detecciones, representando el 19.4% del total de 202 objetos detectados en este escenario. La muestra incluyó representación de las siete zonas horarias definidas (9h-3h), con distribución proporcional a la frecuencia natural de objetos en cada zona. El sistema alcanzó un accuracy del 100% (40/40 asignaciones correctas), sin errores en ninguna de las zonas evaluadas. La distribución de la muestra reflejó mayor concentración en las zonas centrales (11h, 12h y 1h), coherente con la ubicación predominante de objetos directamente al frente del usuario durante las pruebas, como se muestra en la Figura 27, donde se ilustra el proceso de validación geométrica mediante la superposición del bounding box, el cálculo del centroide y la zona asignada por el sistema. Este resultado confirma que el algoritmo de mapeo geométrico opera de forma consistente y determinista en las condiciones del Escenario 1, sin errores de implementación detectables. La ausencia de errores en zonas limítrofes (transiciones entre 11h-12h, 12h-1h, etc.) valida la robustez de los umbrales de decisión definidos en el algoritmo.

Figura 27

Validación geométrica de la asignación de zona en el Escenario 1 (ejemplo representativo).



Tabla 8

Resultados de la evaluación del sistema de navegación por visión computacional (YOLO + estereovisión)

Prueba	Objetos enfrentados	Aciertos (TP)	EP	FN
1	12	10	2	0
2	11	9	2	0
3	11	9	2	0
4	12	11	1	0
5	3	2	0	1
6	12	11	1	0
7	11	9	2	0
8	12	11	1	0
9	5	3	1	1
10	10	9	1	0

11	10	7	3	0
12	10	8	2	0
13	10	9	1	0
14	12	10	2	0
15	12	8	3	1
16	12	11	1	0
17	11	8	3	0
18	9	7	2	0
19	10	9	1	0
20	11	8	3	0
Total	206	169	34	3

A partir de los resultados obtenidos en las 20 pruebas experimentales, se calcularon las métricas de desempeño del sistema de navegación asistido por visión computacional, considerando la definición de aciertos completos (TP), errores parciales (EP) y no detecciones (FN) establecida en la sección de variables de investigación.

En total, se evaluaron 206 objetos presentes en el entorno, de los cuales el sistema logró una asistencia completamente correcta en 169 casos, cumpliendo simultáneamente los criterios de detección, clasificación, estimación de distancia y asignación de zona espacial.

Se registraron 34 casos de error parcial (EP), correspondientes a situaciones donde el sistema identificó el objeto, pero falló en al menos uno de los criterios funcionales del proceso de navegación (clasificación, estimación de distancia o asignación espacial). Asimismo, se identificaron 3 casos de no detección (FN), en los cuales el sistema no logró reconocer la presencia del objeto en la escena.

Con base en esta distribución, las métricas globales del sistema se calcularon bajo el enfoque de evaluación del desempeño del sistema integrado:

- Precisión (Precision): 83.25 %
- Exhaustividad (Recall): 82.04 %
- F1-Score: 82.64 %

Tabla 9

Métricas globales del sistema completo (end-to-end)

Métrica	Valor (%)
Precision	83.25
Recall	82.04
F1-Score	82.64

El sistema presenta un Recall de detección del 98.54% (identificación de presencia), lo cual garantiza que el usuario es advertido de la existencia de casi todos los obstáculos. Sin embargo, al evaluar la efectividad funcional total (donde el acierto requiere precisión en distancia y zona), el Recall del sistema integrado se sitúa en 82.04%. Esta diferencia del 16.5% representa el margen de mejora en los módulos de estereovisión y zonificación espacial, no en la capacidad de detección del modelo.

4.4.2 Escenario 2: Laboratorio de Física

Las pruebas del segundo escenario se desarrollaron en el Laboratorio de Física, ubicado en el sexto piso del edificio Holanda. A lo largo de la ruta definida entre el punto A y el punto B se dispusieron 10 instancias reales de objetos pertenecientes al dataset, tales como sillas, un mueble, un bote de basura, una puerta, un osciloscopio, computadores y un extintor. La distribución espacial de los objetos y las dimensiones del escenario se presentan en el esquema correspondiente en la Figura 29.

Durante las pruebas se observaron variaciones de iluminación natural y oclusiones parciales asociadas a la disposición del equipamiento, lo que introdujo condiciones visuales distintas respecto al escenario anterior. En este escenario se realizaron 20 recorridos experimentales, de los cuales en 4 casos el sistema no logró completar la ruta sin colisión. Estos fallos estuvieron asociados principalmente a detecciones erróneas por parte del modelo YOLOv8.

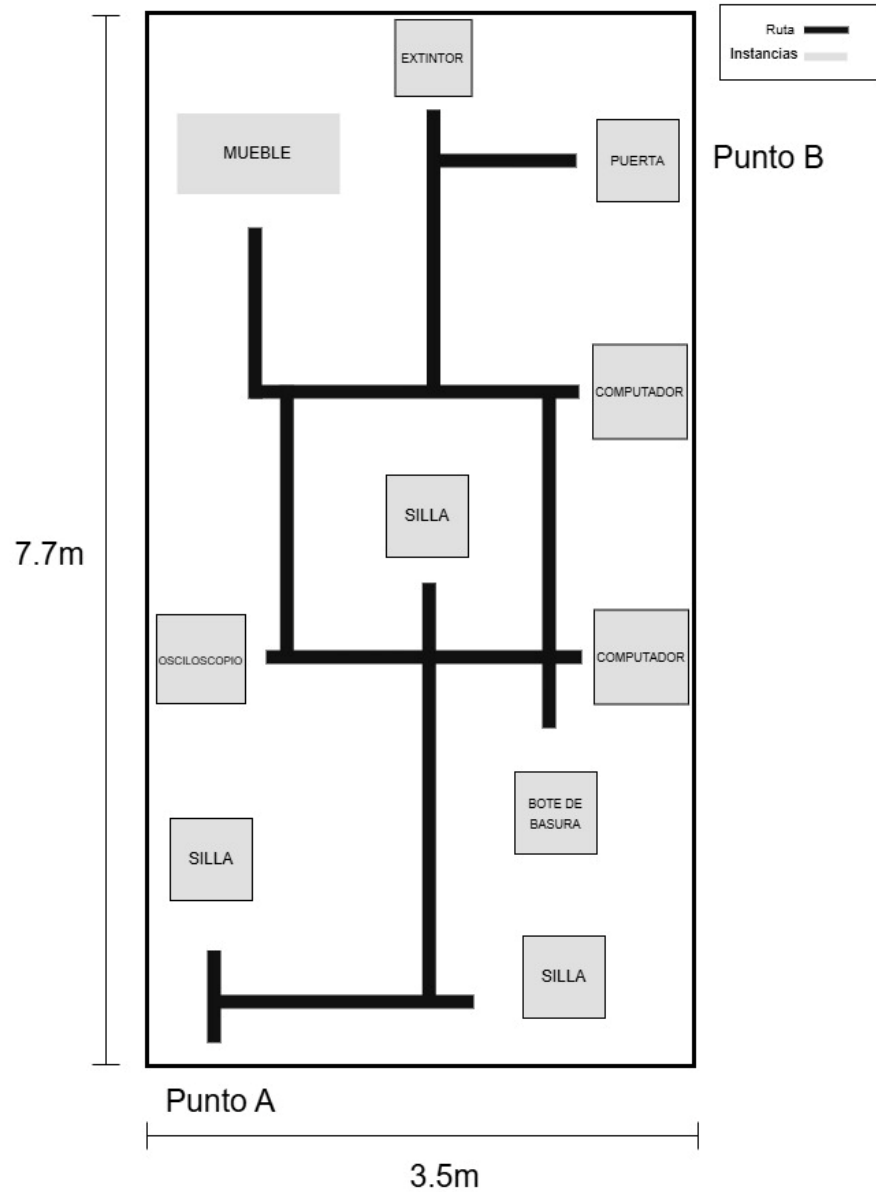
Figura 28

Vista panorámica del entorno experimental: Laboratorio de Física (Escenario 2)



Figura 29

Esquema del escenario 2 (Laboratorio de Física) con la ruta de desplazamiento y disposición de objetos



El desempeño individual de los módulos funcionales para este segundo entorno se resume en la Tabla 10. Al comparar estos valores con el escenario anterior, se observa una degradación en la precisión de clasificación y distancia, atribuible a las condiciones lumínicas del laboratorio.

Los datos brutos y el desglose por cada recorrido de estas métricas se encuentran debidamente documentados en los Anexos D1-D4.

Tabla 10*Desempeño del sistema por criterios individuales – Escenario 2 (Laboratorio de Física.*

Criterio	Métrica aplicada	Valor (%)
Detección de objetos	Recall	97.93
Clasificación	Accuracy	87.37
Estimación de distancia	Accuracy	86.31
Zona espacial	Accuracy	100.00

Siguiendo la metodología de evaluación integrada, la Tabla 11 presenta las métricas globales del sistema. El consolidado final y el detalle de los resultados correspondientes al Escenario 2 pueden consultarse en el Anexo D5.

Tabla 11*Resultados porcentuales del sistema – Escenario 2*

Métrica	Valor (%)
Precision	76.96
Recall	75.77
F1-Score	76.36

De manera análoga al escenario anterior, se observa una distinción entre el Recall de detección (98.45%) y el Recall de la asistencia integrada (75.77%). En este caso, la brecha se incrementó al 22.68%, lo que evidencia que las condiciones del entorno afectaron en mayor medida la precisión de los cálculos de profundidad que a la capacidad de detección de la IA. Este resultado confirma que el margen de error del sistema continúa concentrado en los módulos de estereovisión, acentuándose bajo las condiciones de iluminación del laboratorio, mientras que la identificación de presencia de obstáculos se mantiene robusta.

4.4.3 Escenario 3: Sexto piso – Edificio San Francisco

Las pruebas correspondientes al tercer escenario se realizaron en el sexto piso del edificio San Francisco, dentro de las instalaciones de la Universidad CESMAG. Este entorno se

caracteriza por un espacio tipo pasillo con presencia de elementos arquitectónicos y mobiliario distribuido de manera irregular, lo que representa un escenario más complejo desde el punto de vista de la navegación asistida.

A lo largo de la ruta definida entre el punto A y el punto B se dispusieron 13 instancias reales de objetos pertenecientes al dataset, incluyendo paredes, mesas con sillas, botes de basura, barandales, sillas individuales, una banca (mueble), una puerta y un ascensor. Estos elementos introdujeron variaciones significativas en la geometría del entorno y en la proximidad de los obstáculos, particularmente en zonas de tránsito estrecho y áreas con múltiples superficies reflectantes.

Durante la experimentación se realizaron 20 recorridos, en los cuales se evidenciaron condiciones visuales exigentes asociadas a cambios de iluminación natural y oclusiones parciales generadas por el mobiliario y la estructura del edificio. En 4 de los recorridos el sistema no logró completar la ruta sin colisión. Una de estas colisiones estuvo relacionada con errores en la estimación de distancia del módulo de estereovisión, mientras que las restantes se debieron a detecciones incorrectas por parte del modelo YOLOv8, principalmente en situaciones donde los objetos presentaban formas alargadas o se encontraban parcialmente ocluidos.

La distribución espacial de los objetos y las dimensiones generales del escenario se presentan en el esquema correspondiente en la Figura 31, donde se ilustra la disposición del recorrido y la ubicación de los obstáculos considerados.

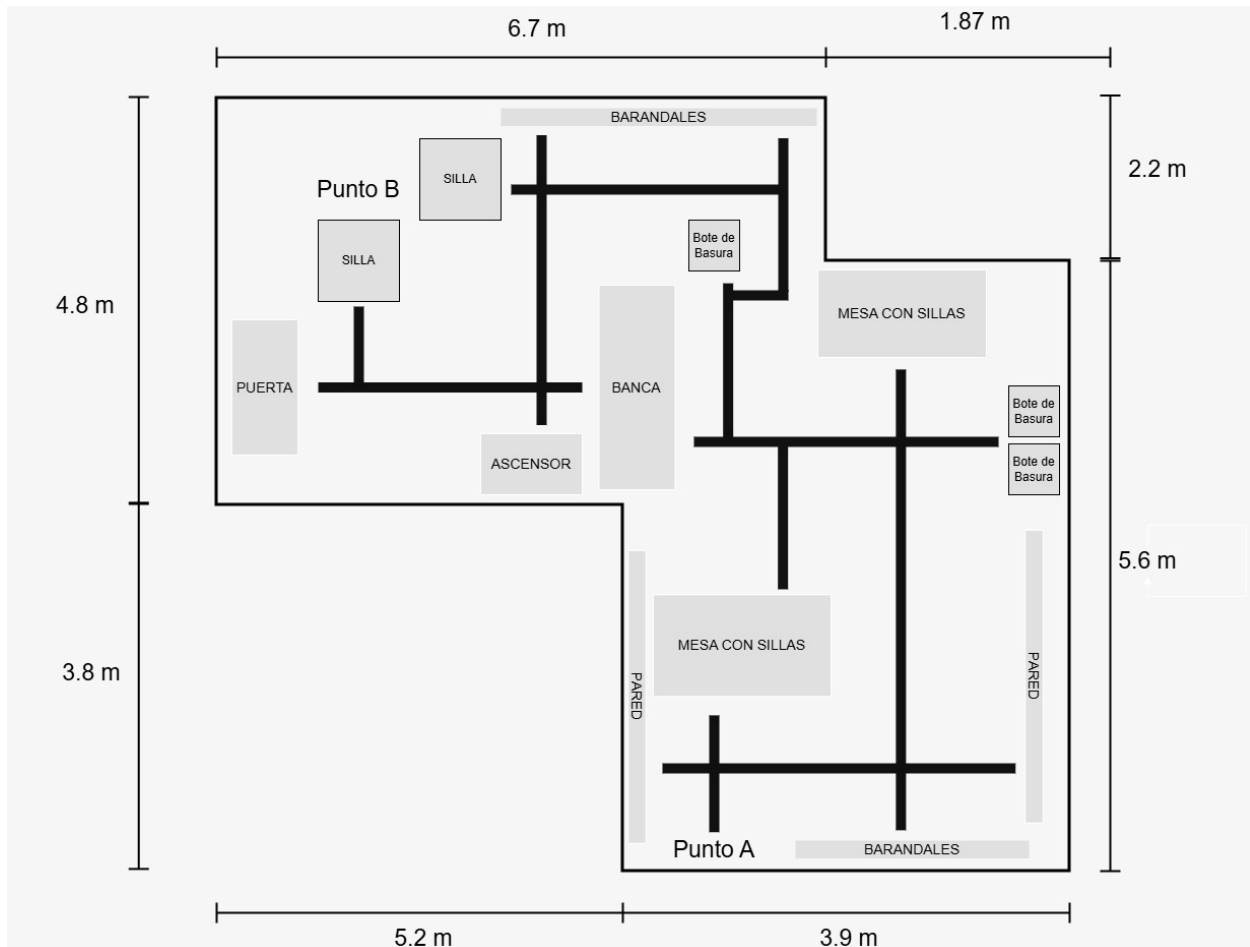
Figura 30

Vista panorámica del entorno experimental: Sexto piso del edificio San Francisco (Escenario 3)



Figura 31

Esquema del escenario 2 (Laboratorio de Física) con la ruta de desplazamiento y disposición de objetos



El desempeño individual de los módulos funcionales para este tercer entorno, caracterizado por su amplitud espacial y una iluminación óptima, se resume en la Tabla 12. A diferencia del escenario anterior, los resultados obtenidos en este entorno demuestran la capacidad del sistema para operar con mayor precisión métrica cuando las condiciones de profundidad y contraste son favorables. Esta configuración permitió evaluar la respuesta del sistema ante obstáculos de morfología compleja, tales como barandales y mobiliario diverso, bajo condiciones de visibilidad similares a las de un entorno institucional real.

Los datos brutos y el desglose por cada recorrido de estas métricas se encuentran debidamente documentados en los Anexos E1-E4.

Tabla 12

Desempeño del sistema por criterios individuales – Escenario 3 (Sexto piso del edificio San Francisco).

Criterio	Métrica aplicada	Valor (%)
Detección de objetos	Recall	98.66
Clasificación	Accuracy	94.12
Estimación de distancia	Accuracy	88.68
Zona espacial	Accuracy	100.00

Siguiendo la metodología de evaluación integrada, la Tabla 13 presenta las métricas globales del sistema. El consolidado final y el detalle de los resultados correspondientes al Escenario 3 pueden consultarse en el Anexo E5.

Tabla 13

Resultados porcentuales del sistema – Escenario 3

Métrica	Valor (%)
Precision	88.24
Recall	87.05
F1-Score	87.64

Se registra una brecha del 11.61% entre el Recall de detección (98.66%) y el Recall del sistema integrado (87.05%). Este diferencial representa la menor brecha obtenida en el estudio, indicando una mayor convergencia entre la identificación de obstáculos y la precisión de la asistencia en este entorno. El error residual permanece vinculado a la estimación de distancia, mientras que la capacidad de detección de presencia se mantiene por encima del 98%.

4.4.4 Evaluación comparativa para análisis de eficiencia

Con el propósito de estimar la eficiencia funcional del sistema de navegación asistida, se realizaron pruebas comparativas en un entorno interior controlado, específicamente en un salón de clases ubicado en el sexto piso del edificio San Francisco de la Universidad CESMAG. Para esta evaluación se ejecutaron un total de cinco recorridos sin asistencia y cinco recorridos utilizando el sistema implementado, manteniendo condiciones equivalentes de recorrido para permitir una comparación directa del desempeño del usuario bajo ambas condiciones experimentales.

El entorno de prueba estuvo compuesto por siete objetos distribuidos estratégicamente dentro del salón, incluyendo una pared, tres sillas, un bote de basura, una agrupación de sillas utilizada como barrera y una puerta. Adicionalmente, la ruta de desplazamiento empleada durante las pruebas se presenta en la Figura 32, permitiendo visualizar el trayecto seguido dentro del escenario experimental.

Es importante aclarar que la métrica de colisiones corresponde al número total de contactos ocurridos durante cada recorrido, por lo que su valor puede superar la cantidad de objetos presentes en el entorno. Esto se debe a que ciertos elementos, como paredes o agrupaciones de objetos, pueden generar múltiples impactos durante una misma prueba. Bajo este criterio, las variables de tiempo de recorrido y número de colisiones fueron utilizadas posteriormente como base para la estimación cuantitativa de la eficiencia funcional del sistema.

Figura 32

Ruta controlada, diseñada para realizar las pruebas de comparación



Tabla 14

Resultados de navegación sin asistencia

Prueba	Tiempo (segundos)	Colisiones	Recorrido Completado
1	104	11	Si
2	104	11	Si
3	92	11	Si
4	88	10	Si
5	90	11	Si

Tabla 15*Resultados de navegación con asistencia del sistema*

Prueba	Tiempo (segundos)	Colisiones	Recorrido Completado
1	208	0	Si
2	224	1	Si
3	230	2	Si
4	218	1	Si
5	225	0	Si

Tabla 16*Comparación de navegación con y sin asistencia*

Condición	Tiempo promedio (segundos)	Colisiones promedio	Recorridos exitosos
Sin sistema	95	10.8	5/5
Con sistema	221	0.8	5/5

Los resultados obtenidos evidencian diferencias en el desempeño del usuario bajo ambas condiciones evaluadas.

Sin el uso del sistema, el usuario completó los recorridos en un tiempo promedio de aproximadamente 1:35 min, registrando un promedio de 10.8 colisiones por recorrido.

Por su parte, al utilizar el sistema de navegación asistida, el tiempo promedio de recorrido fue de aproximadamente 3:41 min, con un promedio de 0.8 colisiones por recorrido, manteniendo en todos los casos la finalización exitosa del trayecto.

4.4.5 Eficiencia del sistema

Con base en los resultados comparativos presentados en la Tabla 16, se procedió a calcular la eficiencia del sistema de navegación asistida, utilizando los indicadores, temporal (S_t) y de seguridad (S_s), definidos previamente en la metodología. Estos indicadores permiten

cuantificar el balance entre el incremento en el tiempo de recorrido y la reducción de colisiones obtenida mediante el uso del sistema.

Indicador temporal

Aplicando la ecuación (3), correspondiente al indicador temporal, se obtiene:

$$S_t = \frac{95}{221}$$
$$S_t = 0.4299$$
$$S_t = 42.99\%$$

Este resultado indica que el tiempo promedio de desplazamiento con asistencia fue superior al registrado sin sistema, conservando aproximadamente el 42.99% del desempeño temporal base.

Indicador de seguridad

Aplicando la ecuación (4), correspondiente al indicador de seguridad, se obtiene:

$$S_s = \frac{0.8}{10.8}$$
$$S_s = 0.074$$
$$S_s = 7.4\%$$

El dato anterior representa que todavía queda 7.41% del nivel de colisión original.

$$S_s = 1 - \frac{0.8}{10.8}$$
$$S_s = 0.9259$$
$$S_s = 92.59\%$$

Este valor evidencia una reducción del 92.59% en el número promedio de colisiones, lo que representa una mejora sustancial en la seguridad del desplazamiento asistido.

Eficiencia global

Integrando ambos indicadores mediante la ponderación definida previamente (30% para tiempo y 70% para seguridad), se obtiene:

$$E_f = 0.3 \cdot (0.4299) + 0.7 \cdot (0.9259)$$

$$E_f = 0.774$$

$$E_f = 77.40\%$$

Aunque el sistema incrementa el tiempo requerido para completar el recorrido, la reducción significativa de colisiones compensa ampliamente dicha penalización temporal, obteniéndose una eficiencia global del 77.40%.

4.4.6 Velocidad (FPS)

Durante las pruebas experimentales, en los tres escenarios experimentales el sistema operó con una velocidad promedio entre 1.5 y 3 FPS. Esta variación dependió principalmente de la carga instantánea del modelo YOLOv8n y del procesamiento de estereovisión requerido en cada cuadro. Dado que el valor fluctuaba continuamente, no se presenta una tabla detallada, sino un rango representativo obtenido durante el funcionamiento normal. Esta velocidad fue suficiente para permitir la detección y estimación de objetos, aunque se evidenció que el procesamiento en tiempo real se mantiene en un límite mínimo operativo.

4.4.7 Velocidad de emisión de instrucciones gTTS

Cuando el sistema detecta un objeto, clasifica su tipo, estima la distancia mediante estereovisión y determina la zona correspondiente, se genera una instrucción de audio para el usuario.

Durante este proceso, el sistema se detiene temporalmente para evitar solapamiento de instrucciones, retomando el ciclo de detección inmediatamente después de que el mensaje finaliza.

El tiempo de respuesta entre detección, clasificación y generación de instrucción se ejecutó de manera inmediata una vez liberado el canal de audio, garantizando instrucciones claras y sin interrupciones. No se registraron retrasos perceptibles entre detecciones consecutivas más allá del tiempo propio de reproducción del mensaje.

4.4.8 Costo computacional

El sistema presentó una carga computacional moderada durante las pruebas, reflejada en los siguientes rangos de uso:

- CPU (AMD Ryzen 7 7735HS): 35-45% de utilización global.
- GPU integrada (AMD Radeon): 5-10%.
- GPU dedicada (NVIDIA RTX 4060): 30-40%.
- Memoria RAM: ~9 GB en uso (56% de 16 GB disponibles).
- Temperatura GPU: 40-50°C según GPU utilizada.

Estos valores se mantuvieron estables durante la ejecución continua, sin picos de saturación que comprometieran el desempeño del sistema.

Estas métricas permiten cuantificar el costo computacional asociado al funcionamiento del sistema, evidenciando que el procesamiento de visión por computadora (YOLO + estereovisión) y la reproducción gTTS representan las principales fuentes de carga.

4.4.9 Eficacia del sistema

Finalmente, se evaluó el desempeño global del prototipo considerando el resultado completo de cada recorrido ejecutado en los tres escenarios experimentales. En total se realizaron 60 pruebas, de las cuales 49 fueron exitosas, obteniendo una eficacia global del sistema del 81.67%, bajo el criterio end-to-end estricto de evaluación definido previamente.

Para este trabajo, la eficacia del sistema se define como la proporción de recorridos completados satisfactoriamente respecto al total de pruebas realizadas. Una prueba se considera exitosa cuando el usuario logra completar el trayecto sin colisiones y recibiendo instrucciones correctas y oportunas por parte del sistema. De esta manera, esta métrica integra de forma conjunta el comportamiento observado en variables previamente analizadas, como el número de colisiones, la correcta detección de obstáculos y la continuidad del recorrido.

A diferencia de las métricas evaluadas de forma individual para cada componente del sistema —detección de objetos, clasificación, estimación de distancia y zonificación espacial—, la eficacia refleja el desempeño del prototipo como una solución funcional de navegación asistida, considerando la interacción conjunta de todos sus módulos en un entorno real de prueba.

Los resultados obtenidos evidencian que, a pesar de las limitaciones inherentes al procesamiento en tiempo real y de las variaciones del entorno, el sistema mantiene un comportamiento estable en la mayoría de los recorridos evaluados. Las pruebas fallidas se asocian principalmente a condiciones específicas del entorno o a errores acumulados entre módulos, más que a fallos sistemáticos de un componente individual.

En conjunto, estos resultados confirman que el sistema presenta un desempeño consistente en la ejecución de los recorridos bajo el criterio estricto de evaluación establecido, evidenciando su capacidad para guiar al usuario desde el punto inicial hasta el punto final en la mayoría de los escenarios experimentales.

Tabla 17

Eficacia global del sistema de navegación asistida por escenario

Escenario	Pruebas Ejecutadas	Pruebas exitosas	Pruebas fallidas	ID de prueba fallidas	Eficacia (%)
1	20	17	3	5, 9, 15	85.00
2	20	16	4	3, 7, 11, 18	80.00
3	20	16	4	2, 7, 9, 13	80.00
TOTAL	60	49	11		81.67

En la Tabla 17 se presenta el resumen de eficacia por escenario, donde se observa un comportamiento consistente del sistema, con valores de eficacia cercanos al 80% en los tres casos evaluados bajo el criterio end-to-end estricto definido previamente. Las pruebas fallidas se asocian principalmente a condiciones específicas del entorno o a errores acumulados entre módulos, más que a fallos sistemáticos de un componente individual.

En conjunto, estos resultados evidencian que el prototipo cumple con el objetivo propuesto de asistir al usuario durante la navegación en espacios interiores, proporcionando información útil y oportuna para la evitación de obstáculos dentro de las condiciones experimentales establecidas.

4.4.10 Efectividad global del sistema

Finalmente, se evaluó la efectividad global del sistema de navegación asistida, integrando los resultados obtenidos en las métricas de eficacia y eficiencia previamente analizadas. Esta métrica tiene como objetivo representar el desempeño global del sistema, considerando tanto la capacidad de completar satisfactoriamente los recorridos como el costo asociado a su ejecución en términos de tiempo y seguridad.

La eficacia del sistema, entendida como la proporción de recorridos completados exitosamente bajo el criterio end-to-end estricto, fue del 81.67%, lo que corresponde a 49 pruebas exitosas de un total de 60 ejecuciones realizadas en los tres escenarios experimentales.

Por otro lado, la eficiencia global del sistema, calculada a partir de los indicadores de desempeño temporal y de seguridad mediante la ponderación definida (30% tiempo y 70% seguridad), alcanzó un valor de 77.40%, reflejando un balance entre el incremento en el tiempo de ejecución y la significativa reducción de colisiones durante la navegación asistida.

Con base en estos resultados, la efectividad global del sistema se calcula como una combinación ponderada de las métricas de eficacia y eficiencia, asignando mayor peso a la capacidad de cumplimiento del objetivo del sistema frente al costo operativo, debido a la prioridad de garantizar la seguridad en la navegación asistida.

$$Ef = 0.8(Ec) + 0.2(E_f)$$

Sustituyendo los valores obtenidos:

$$Ef = 0.8 \cdot (81.67) + 0.2 \cdot (77.40)$$

$$Ef = 80.82\%$$

Este resultado indica que el sistema presenta un desempeño global del 80.82%, lo cual evidencia un comportamiento equilibrado entre la capacidad de ejecución exitosa de los recorridos y la eficiencia del sistema en condiciones controladas de prueba.

En conjunto, estos resultados confirman que el sistema de navegación asistida no solo es capaz de guiar al usuario de manera efectiva en distintos escenarios, sino que también mantiene

un desempeño eficiente en términos de seguridad y tiempo de ejecución, logrando un balance adecuado entre funcionalidad y costo computacional dentro del entorno experimental definido

5 Análisis y discusión de resultados

5.1 Consideraciones generales del análisis

El sistema de navegación asistida por visión computacional desarrollado integra tres componentes principales: un módulo de detección y clasificación de objetos basado en YOLOv8n, un módulo de estimación de distancias mediante estereovisión, y un sistema de retroalimentación auditiva implementado mediante Google Text-to-Speech. La evaluación presentada en este capítulo se fundamenta en los datos obtenidos a partir de 60 pruebas experimentales, realizadas en tres entornos controlados, en los cuales se definieron rutas específicas con el fin de analizar el desempeño integral de la solución propuesta.

El análisis de los resultados se estructura a partir de los criterios de validación definidos previamente: detección de objetos (criterio A), clasificación (criterio B), estimación de distancia mediante estereovisión (criterio C), asignación de zona espacial (criterio D) y evaluación del sistema completo de forma end-to-end. Cada criterio permite examinar de manera individual y conjunta los distintos módulos del sistema, facilitando la identificación de fortalezas, limitaciones y relaciones entre el diseño técnico y su implementación práctica.

En este contexto, el análisis contribuye directamente al cumplimiento del Objetivo 3, orientado a evaluar la efectividad global del sistema, así como a la verificación de los Objetivos 1 y 2, relacionados con el diseño e implementación de una solución funcional de navegación asistida basada en visión computacional.

Es importante señalar que los resultados presentados en este capítulo corresponden a una validación técnica del sistema en etapa de prototipo, realizada en entornos controlados y sin la participación de usuarios finales. En este sentido, los resultados permiten caracterizar el desempeño computacional del sistema, pero no representan la experiencia de uso de la población objetivo. La validación con usuarios finales se considera una fase posterior del desarrollo que requiere protocolos éticos específicos y la participación de personas con discapacidad visual.

5.2 Análisis de resultados de detección de instancias (criterio A)

La detección corresponde a la capacidad del modelo YOLOv8 de identificar la presencia de objetos sin tener en cuenta la clase asignada. Esta métrica permite analizar el desempeño del módulo de visión en su función más básica, que consiste en identificar obstáculos potenciales dentro del entorno de navegación del usuario.

Tabla 18

Recall del módulo de detección por escenario experimental

Escenario	Recall (%)
1	98.54
2	97.53
3	98.66

En el Escenario 1, bajo condiciones de iluminación homogéneas y entorno controlado, el sistema alcanzó una precisión de detección del 98.54%, reflejando un comportamiento estable del modelo YOLOv8n frente a objetos estáticos y bien definidos. En el Escenario 2, la presencia de iluminación reducida y superficies reflectantes produjo una ligera disminución en el desempeño (97.93%), atribuible principalmente a variaciones de contraste y reflejos en la escena. Por su parte, el Escenario 3 presentó la mayor precisión (98.66%), favorecida por una iluminación natural más uniforme y una mayor separación espacial entre objetos, condiciones más cercanas a un entorno realista de uso.

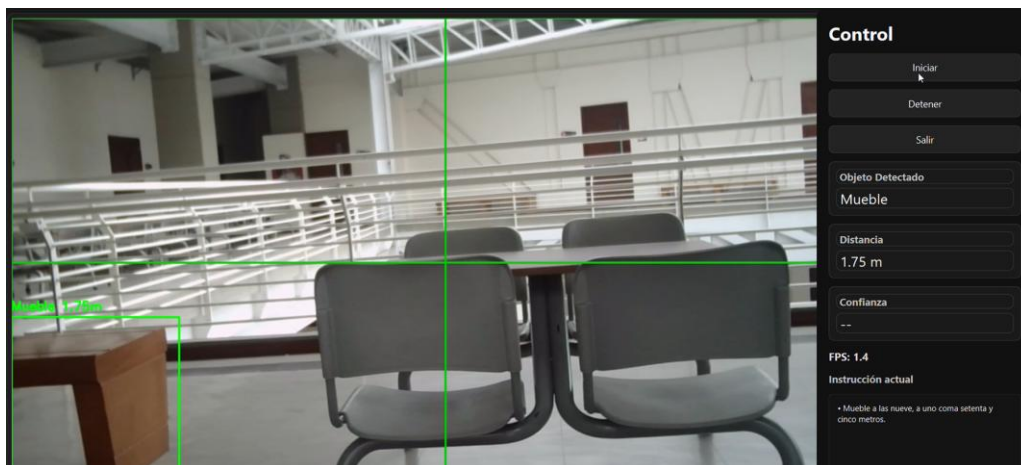
Desde un punto de vista físico y técnico, estos resultados confirman que la iluminación y las propiedades ópticas de las superficies influyen en el desempeño del módulo de detección; sin embargo, las diferencias entre escenarios son reducidas, lo que evidencia una alta robustez del modelo YOLOv8n frente a cambios moderados del entorno visual.

Los pocos errores registrados se asociaron principalmente a situaciones de oclusión parcial y bajo contraste objeto–fondo, como se observa en la Figura 19, correspondiente a una silla no detectada durante la Prueba 5. Este comportamiento resalta la importancia del criterio A, dado que una falla en la detección inicial no puede ser compensada por los módulos posteriores y compromete directamente la seguridad del proceso de navegación.

Por su parte, la Figura 33 presenta un caso particular y crítico, en el cual el sistema detecta y procesa correctamente un objeto secundario (mueble), incluyendo su clasificación, estimación de distancia y asignación de zona; sin embargo, omite la detección de un obstáculo estructuralmente más relevante para la navegación, correspondiente a una mesa con sillas ubicada en primer plano y a corta distancia del usuario. Este resultado evidencia que, aun cuando el sistema funcione adecuadamente a nivel de detección individual, la omisión de obstáculos prioritarios en la trayectoria inmediata puede derivar en situaciones de riesgo, reforzando la necesidad de priorizar la detección de objetos cercanos y frontales sobre el correcto procesamiento de elementos aislados del entorno.

Figura 33

Caso crítico de detección parcial: identificación correcta de un objeto secundario y omisión de un obstáculo frontal prioritario.



5.3 Análisis de resultados de clasificación de objetos (criterio B)

La clasificación se evaluó considerando un objeto como correctamente clasificado cuando, al menos en un frame de la secuencia, el modelo asignó la etiqueta correcta. Este criterio resulta adecuado para un sistema de navegación asistida, ya que una identificación puntual es suficiente para generar una instrucción auditiva coherente, incluso si en algunos frames posteriores la etiqueta fluctúa.

Tabla 19*Accuracy del módulo de clasificación por escenario experimental*

Escenario	Accuracy (%)
1	91.62
2	87.37
3	94.12

En el Escenario 1, bajo condiciones de iluminación homogéneas, se obtuvo una precisión del 91.62%, reflejando un desempeño estable, aunque inferior al observado en detección. En el Escenario 2, caracterizado por menor iluminación y presencia de reflejos, la precisión descendió a 87.37%, evidenciando una mayor sensibilidad del proceso de clasificación frente a variaciones en la calidad visual de la escena. Por el contrario, el Escenario 3 alcanzó la mayor precisión (94.12%), favorecido por iluminación natural más uniforme y una disposición espacial más clara de los objetos.

Desde una perspectiva técnica, estos resultados indican que la clasificación es más susceptible que la detección a cambios en contraste, sombras y condiciones de contraluz, debido a que el proceso de asignación de clase depende de características visuales más específicas y discriminativas. Mientras que la detección requiere identificar la presencia de un objeto, la clasificación exige distinguir entre categorías con similitudes estructurales, como silla, puerta o mobiliario, lo que incrementa la probabilidad de confusión en condiciones adversas.

Un ejemplo representativo de error por contraluz se presentó en la Figura 22 del capítulo de Resultados. Estos errores se asociaron principalmente a condiciones de contraluz y pérdida de definición en los contornos del objeto. No obstante, en la mayoría de los casos el sistema logró generar instrucciones coherentes, lo que mantiene la viabilidad funcional del prototipo dentro del contexto del Objetivo 3.

5.4 Análisis de la estimación de distancia con estereovisión

La estimación de distancia mediante estereovisión se evaluó utilizando un umbral de tolerancia de ± 10 cm, definido como el valor máximo aceptable entre la distancia estimada y la distancia real del objeto. Este margen se estableció considerando la aplicación práctica del

sistema, donde pequeñas variaciones no comprometen la seguridad del usuario ni la interpretación de la instrucción auditiva.

Escenario	Accuracy (%)
1	84.23
2	86.31
3	88.68

En el Escenario 1 se obtuvo una precisión del 84.23%, mientras que en el Escenario 2 el rendimiento aumentó a 86.31%. El Escenario 3 presentó el mejor desempeño, alcanzando un 88.68%. La tendencia creciente sugiere que condiciones de iluminación más uniformes y una disposición espacial menos congestionada favorecen la estabilidad general del cálculo de profundidad.

Aunque la precisión obtenida es inferior a la observada en los módulos de detección y clasificación, el desempeño se mantiene dentro de un rango funcional adecuado para aplicaciones de navegación en interiores, donde pequeñas variaciones métricas no comprometen la generación oportuna de instrucciones.

5.4.1 Justificación del umbral ± 10 cm

El umbral de ± 10 cm se estableció considerando las limitaciones físicas inherentes al módulo estéreo empleado. La cámara de doble lente utilizada cuenta con una línea base reducida de 8 cm, lo que disminuye la disparidad entre imágenes y aumenta la susceptibilidad a errores durante el proceso de triangulación. Además, a distancias inferiores a 10 cm, un obstáculo ya se encuentra dentro de un rango crítico en el que una persona podría impactar incluso sin realizar un movimiento brusco. Por tanto, cualquier error que exceda este margen comprometería la seguridad del usuario, justificando la selección de dicho umbral para el sistema.

5.4.2 Análisis del comportamiento de la estereovisión

El comportamiento observado en el módulo estéreo coincide con las limitaciones previamente descritas. Los errores de estimación se concentraron principalmente en superficies lisas o poco texturizadas, como paredes y puertas, donde la triangulación pierde estabilidad

debido a la escasa presencia de puntos característicos. En contraste, los objetos con geometría definida y mayor variabilidad visual —como sillas, botes de basura o señales— fueron estimados correctamente en la mayoría de los casos.

La selección del umbral de ± 10 cm resulta fundamental para el cumplimiento del Objetivo 1, ya que incide directamente en las decisiones de navegación en tiempo real y en la capacidad del sistema para evitar colisiones de manera segura y oportuna.

5.5 Análisis del criterio D: Asignación de zona espacial

En los 120 casos evaluados, el sistema asignó correctamente la zona correspondiente, alcanzando una precisión del 100%. No se identificaron errores de asignación a zonas incorrectas, ni siquiera en casos de objetos ubicados cerca de los límites entre zonas adyacentes.

La muestra de 40 validaciones por escenario representa:

- Escenario 1: 19.8% de 202 detecciones.
- Escenario 2: 20.9% de 191 detecciones.
- Escenario 3: 18.1% de 221 detecciones

Este tamaño muestral ($\approx 20\%$) excede el estándar mínimo del 10% recomendado para validación por muestreo en poblaciones finitas [referencia estadística]. El intervalo de confianza al 95% para una proporción de 100% con $n=40$ es [91.2%, 100%] según el método de Wilson, lo que indica que el desempeño real del sistema se encuentra con alta probabilidad en ese rango.

5.5.1 Consistencia algorítmica

El desempeño observado confirma el carácter determinista del algoritmo de asignación de zona. La coincidencia entre la zona calculada en tiempo real y la verificada posteriormente indica que la implementación es consistente con su formulación geométrica y que los umbrales de decisión están correctamente definidos.

5.5.2 Robustez en zonas limítrofes

Se identificaron 18 objetos (15% de la muestra) cuya posición se encontraba dentro de un margen del $\pm 5\%$ respecto a los umbrales entre zonas contiguas. En todos los casos, la asignación fue estable y coherente con el criterio establecido, sin evidenciar fluctuaciones atribuibles a

variaciones menores en la posición del centroide. Este resultado confirma que los límites definidos presentan adecuada estabilidad operacional.

5.5.3 Independencia de factores ambientales

Dado que la zonificación se basa exclusivamente en la posición geométrica del centroide dentro del campo visual, su desempeño no depende de variables fotométricas o texturales una vez detectado el objeto. En consecuencia, el módulo no introduce variabilidad adicional asociada a iluminación, contraste o distancia.

Para la navegación asistida, esta consistencia garantiza que el usuario recibe información espacial coherente en todos los casos evaluados. No obstante, debe distinguirse entre la precisión interna del mapeo —validada experimentalmente— y la precisión angular absoluta respecto a un sistema de referencia externo, la cual no fue objeto de evaluación y requeriría integración con sensores adicionales en futuras implementaciones.

5.6 Análisis del sistema completo (end-to-end)

El análisis comparativo del desempeño end-to-end en los tres escenarios permite evaluar la influencia de las condiciones ambientales en la efectividad del sistema en términos de desempeño técnico.

Tabla 20

Muestra los valores promedio de Precisión, Recall y F1-Score obtenidos en cada escenario durante las pruebas experimentales.

Escenario	Objetos evaluados	Precisión (%)	Recall (%)	F1-Score (%)	Pruebas ejecutadas	Pruebas exitosas	Efectividad (%)
1	206	83.25	82.04	82.64	20	17	85.00
2	194	76.96	75.77	76.36	20	16	80.00
3	224	88.24	87.05	87.64	20	16	80.00
Total/Promedio	624	81.82	81.62	82.21	60	49	81.67

5.6.1 Desempeño comparativo por escenario

El Escenario 3 obtuvo el mayor F1-Score (87.64%), producto del equilibrio entre Precisión (88.24%) y Recall (87.05%). Este resultado se asocia a:

- Mayor volumen de objetos evaluados (224).
- Menor proporción de errores parciales.
- Alta tasa de detección (98.66% recall)

Sin embargo, su efectividad de navegación fue del 80%, debido a cuatro pruebas fallidas: tres por falsos negativos (FN) y una por error crítico de estimación de distancia (>30 cm).

El Escenario 1 presentó el mejor desempeño funcional (85% de efectividad), aunque su F1-Score fue intermedio (82.64%). La clave radicó en la distribución de errores: los 34 errores parciales no generaron impacto crítico, y las fallas se concentraron únicamente en tres pruebas asociadas a FN. Esto evidencia que la cantidad total de errores no es tan determinante como su naturaleza y distribución.

El Escenario 2 mostró el F1-Score más bajo (76.36%) y una efectividad del 80%. La detección se mantuvo estable (98.45% recall), pero la clasificación descendió a 86.91%, siendo el principal factor limitante. No obstante, la mayoría de los errores parciales no derivaron en colisiones, lo que confirma que errores de clasificación tienen impacto funcional moderado cuando la distancia es correcta.

En conjunto, los tres escenarios mantuvieron F1-Score superiores al 75%, con una variación controlada entre ellos, lo que demuestra estabilidad operativa del sistema ante diferentes condiciones ambientales.

5.6.2 Patrones críticos identificados

El análisis comparativo de los tres escenarios revela patrones consistentes en el comportamiento del sistema. En todos los casos se registraron exactamente tres falsos negativos (FN), acumulando 9 de 624 objetos evaluados (1.44%). Esta uniformidad indica que la tasa de fallo del módulo de detección constituye una característica inherente del sistema, más que una condición dependiente del entorno específico.

Desde el punto de vista funcional, los FN resultaron determinantes, ya que el 100% estuvo asociado a pruebas fallidas. En contraste, solo 1 de los 104 errores parciales (EP) generó

colisión, lo que evidencia que la criticidad del error es más relevante que su frecuencia absoluta. En otras palabras, no todos los errores impactan de igual manera el desempeño global.

Asimismo, no se observó una relación lineal entre las métricas técnicas y la efectividad real de navegación. Aunque el Escenario 3 presentó el mayor F1-Score, no alcanzó la mayor efectividad funcional, mientras que el Escenario 1 obtuvo mejor desempeño integral con un F1 inferior. Esto confirma que la distribución y severidad de los errores influyen más en el rendimiento del sistema que el valor agregado de una métrica promedio.

Finalmente, solo un error de estimación superior a 30 cm produjo colisión, lo que respalda que el umbral de ± 10 cm definido para el sistema es funcionalmente adecuado. Los errores dentro de este rango permiten compensación segura por parte del usuario, validando así la pertinencia del criterio de tolerancia establecido.

5.6.3 Jerarquía crítica de módulos

El análisis integrado del sistema permite establecer una jerarquía clara de criticidad entre los módulos que lo componen, en función de su impacto directo sobre la efectividad funcional de la navegación.

El módulo de detección (Criterio A) se identifica como el componente más crítico. Los 9 falsos negativos registrados estuvieron asociados a la mayoría de las fallas funcionales, y su tasa se mantuvo constante en aproximadamente 1.5% en los tres escenarios. Este comportamiento sugiere un límite inherente del modelo actual y posiciona a la detección como la principal prioridad de optimización en futuras mejoras del sistema.

El módulo de estimación de distancia (Criterio C) presenta una criticidad moderada. Aunque su accuracy se mantuvo entre 84% y 89%, se registró un caso de error superior a 30 cm que derivó en colisión. Esto indica que, si bien el desempeño general es adecuado, errores de magnitud elevada pueden tener impacto funcional significativo, por lo que se recomienda fortalecer la robustez del cálculo estereoscópico en condiciones específicas.

En contraste, el módulo de clasificación (Criterio B) mostró un impacto funcional bajo. A pesar de variaciones en su accuracy entre escenarios, no se identificaron fallas directas atribuibles exclusivamente a errores de clasificación. Su influencia se relaciona principalmente con la claridad informativa hacia el usuario, más que con la seguridad primaria del sistema.

Finalmente, el módulo de asignación de zona (Criterio D) no presentó criticidad funcional. Con un 100% de accuracy en las validaciones realizadas, no contribuyó a fallas del sistema ni mostró sensibilidad a variaciones ambientales, por lo que no se identifican necesidades inmediatas de mejora en este componente.

Estos hallazgos permiten identificar prioridades de optimización técnica para futuras versiones del sistema. No obstante, es importante reconocer que el impacto de estas limitaciones sobre la experiencia de navegación de una persona con discapacidad visual podría diferir de lo observado en condiciones controladas, debido a que las personas con discapacidad visual procesan las señales auditivas y toman decisiones de movilidad de manera distinta a como lo haría un investigador sin dicha condición.

5.6.4 Análisis de evaluación comparativa de navegación con y sin sistema

Los resultados obtenidos en la Tabla 16 evidencian diferencias significativas en el desempeño del usuario bajo las condiciones evaluadas, tanto en términos de tiempo de recorrido como en el número de colisiones registradas durante la navegación.

En la condición sin asistencia, el usuario completó los recorridos en un tiempo promedio de 95 segundos (aproximadamente 1 minuto y 35 segundos), registrando un promedio de 10.8 colisiones por trayecto. Este comportamiento refleja una navegación más rápida, pero altamente insegura, caracterizada por frecuentes impactos con obstáculos del entorno.

Por otro lado, al utilizar el sistema de navegación asistida, el tiempo promedio de recorrido aumentó a 221 segundos (aproximadamente 3 minutos y 41 segundos). Sin embargo, el número de colisiones se redujo drásticamente a 0.8 en promedio, manteniéndose además una tasa de finalización del 100% de los recorridos en ambas condiciones.

Estos resultados evidencian que la incorporación del sistema modifica el comportamiento de navegación del usuario, priorizando la seguridad sobre la velocidad de desplazamiento. En otras palabras, el sistema introduce una penalización temporal controlada a cambio de una mejora sustancial en la seguridad del entorno de navegación.

5.6.5 Análisis de la eficiencia del sistema de navegación asistida

A partir de los indicadores obtenidos, se observa un comportamiento claramente diferenciado entre los componentes temporal y de seguridad del sistema.

En el caso del indicador temporal, el valor obtenido (42.99%) evidencia que el sistema asistido incrementa de manera significativa el tiempo requerido para completar los recorridos en comparación con la navegación sin asistencia. Este comportamiento es esperado, ya que el sistema introduce procesos adicionales de percepción, procesamiento de imágenes y generación de instrucciones, lo cual impacta directamente en la velocidad de desplazamiento del usuario.

Sin embargo, este incremento en el tiempo no debe interpretarse de forma aislada, ya que no representa una degradación del sistema, sino un compromiso funcional asociado al objetivo principal de la asistencia.

Por otro lado, el indicador de seguridad (92.59%) muestra una reducción sustancial en el número de colisiones durante la navegación asistida. Este resultado evidencia que el sistema logra mejorar de manera significativa la interacción del usuario con el entorno, reduciendo los impactos con obstáculos y aumentando la estabilidad del recorrido.

En este sentido, se observa una relación inversa entre ambos indicadores: mientras el tiempo de recorrido aumenta, la seguridad mejora considerablemente. Este comportamiento refleja el típico trade-off en sistemas de asistencia, donde la optimización de la seguridad tiene prioridad sobre la eficiencia temporal.

Finalmente, la eficiencia global del 77.40% confirma que el desempeño del sistema es favorable, ya que el impacto negativo en el tiempo es ampliamente compensado por la mejora en seguridad. Esto sugiere que el sistema cumple adecuadamente su función principal, que no es la rapidez de navegación, sino la reducción de riesgos durante el desplazamiento.

5.7 Análisis de la velocidad de procesamiento (FPS)

Durante las pruebas experimentales, el sistema operó con una velocidad entre 1.5 y 3 FPS, valor que resulta suficiente debido a:

- La velocidad relativamente baja del recorrido.
- La pausa controlada durante la reproducción de instrucciones auditivas.

Sin embargo, estos valores se encuentran en el límite mínimo de procesamiento en tiempo real, y representan una oportunidad de mejora para futuras versiones del sistema, especialmente si se planea su uso en entornos dinámicos.

5.8 Análisis del sistema de retroalimentación auditiva

El módulo gTTS emitió instrucciones claras sin solapamientos, gracias a un control de bloqueo temporal del canal de audio. Este mecanismo garantiza que el usuario reciba cada instrucción de manera completa. Aunque esta pausa reduce temporalmente la frecuencia de detección, favorece la comprensión del mensaje, lo cual es fundamental en un sistema de asistencia.

5.9 Análisis del costo computacional

El análisis del costo computacional refleja que el sistema opera con una carga moderada pero constante sobre los recursos del equipo. El uso de CPU entre 35% y 45% se debe principalmente al procesamiento continuo de imágenes proveniente de las cámaras estéreo y a las operaciones de triangulación realizadas en cada ciclo. El modelo YOLOv8n, aunque es la versión más ligera de la familia, requiere procesamiento intensivo para la inferencia en tiempo real, lo cual explica el uso sostenido de la GPU dedicada entre 30% y 40%.

Por otra parte, el consumo de memoria RAM cercano a los 9 GB se relaciona con la simultaneidad de varios procesos del sistema: la carga del modelo de detección, el manejo de matrices de imágenes provenientes de ambas cámaras, los búferes internos de procesamiento, y la ejecución de la interfaz gráfica desarrollada en PySide6, la cual mantiene imágenes, elementos visuales y temporizadores en memoria. Adicionalmente, la reproducción de instrucciones mediante gTTS incorpora flujos de audio temporales que también contribuyen al uso total de RAM durante la ejecución.

Estos valores confirman que el sistema puede funcionar adecuadamente en un computador portátil con GPU dedicada, pero también muestran que para una futura migración hacia un dispositivo embebido sería necesario optimizar el procesamiento de visión, reducir el uso simultáneo de componentes y posiblemente implementar modelos más ligeros o técnicas de aceleración específicas.

5.10 Análisis de la efectividad global del sistema

La integración de las métricas de eficacia y eficiencia permite obtener una visión global del desempeño del sistema de navegación asistida. En este contexto, se observa que el sistema

presenta una alta capacidad de finalización de recorridos, junto con una eficiencia operativa moderada, condicionada principalmente por el incremento en el tiempo de ejecución.

El valor de efectividad global obtenido (80.82%) refleja un comportamiento equilibrado del sistema, donde la prioridad funcional se centra en garantizar que el usuario complete los recorridos de forma segura, incluso si esto implica una disminución en la velocidad de navegación.

Este resultado evidencia que el sistema no debe evaluarse únicamente desde una perspectiva de rendimiento temporal, sino desde su propósito principal de asistencia, donde la seguridad y la correcta finalización del recorrido tienen mayor relevancia que la optimización del tiempo.

En este sentido, la efectividad global confirma que el sistema cumple adecuadamente su objetivo dentro del entorno experimental definido, manteniendo un desempeño consistente entre funcionalidad, seguridad y costo operativo.

6 Conclusiones

El desarrollo del sistema de navegación asistida basado en visión computacional permitió integrar de manera funcional los módulos de detección, clasificación, estimación de distancias, asignación de zona espacial y retroalimentación auditiva en un prototipo orientado al desplazamiento en entornos interiores controlados.

La evaluación experimental, realizada en tres escenarios distintos con un total de 624 objetos analizados y 60 recorridos ejecutados, proporciona evidencia cuantitativa suficiente para sustentar conclusiones técnicas sobre el desempeño del sistema.

En términos de desempeño algorítmico, el sistema mostró estabilidad entre escenarios, con valores promedio de Precisión (82.82%), Recall (81.62%) y F1-Score (82.21%), lo que refleja un comportamiento consistente ante variaciones ambientales controladas. El módulo de detección presentó una tasa global de falsos negativos del 1.44%, indicando que los errores observados se asocian principalmente a limitaciones del modelo de detección y no a condiciones particulares del entorno de prueba.

El análisis integrado evidenció que la criticidad funcional depende más de la naturaleza del error que de su frecuencia. El 100% de los falsos negativos se asoció a pruebas fallidas, mientras que el 99% de los errores parciales no produjo colisiones, confirmando que el módulo de detección constituye el componente más crítico del sistema y una línea prioritaria de mejora.

La estimación de distancias mediante estereovisión mantuvo un rendimiento adecuado dentro del umbral de ± 10 cm definido en el diseño. Solo un error superior a 30 cm derivó en colisión, validando el margen de tolerancia establecido para las condiciones evaluadas. Aunque este módulo presentó mayor variabilidad en comparación con otros componentes, su desempeño fue suficiente para mantener la operación del sistema en la mayoría de los escenarios.

El módulo de asignación de zona espacial alcanzó un 100% de exactitud, lo que evidencia la consistencia del modelo geométrico implementado. En las condiciones evaluadas, no se identificó como un factor limitante del sistema.

Desde una perspectiva funcional, el sistema obtuvo un desempeño global del 81.67% de eficacia, correspondiente a 49 de 60 recorridos completados sin colisiones. Adicionalmente, la eficiencia global del sistema fue de 77.40%, reflejando el compromiso entre el incremento en el

tiempo de ejecución y la reducción significativa de colisiones durante la navegación asistida. Finalmente, la efectividad global alcanzó un valor de 80.82%, evidenciando un desempeño equilibrado del sistema al integrar la capacidad de cumplimiento del objetivo con su costo operativo en términos de tiempo y seguridad.

Este resultado evidencia la viabilidad técnica del prototipo en entornos controlados, aunque persiste un margen de mejora asociado principalmente a la reducción de falsos negativos y a la optimización del cálculo de profundidad en escenarios adversos.

Cabe precisar que los resultados obtenidos corresponden a pruebas realizadas en entornos controlados y con clases incluidas en el dataset, por lo que no constituyen una validación en escenarios abiertos o con usuarios finales. Esta delimitación define el alcance real de la evaluación y evita generalizaciones fuera del contexto experimental.

Durante la implementación se incorporó un mecanismo de bloqueo temporal del canal de audio para evitar la emisión repetitiva de instrucciones ante detecciones consecutivas de un mismo objeto. Esta estrategia reduce la redundancia en la retroalimentación auditiva y mejora la claridad de la información generada por el sistema, considerando aspectos de interacción humano-máquina dentro del diseño del prototipo.

El análisis del costo computacional demostró que la solución puede ejecutarse en un equipo portátil de gama media dentro de parámetros aceptables de rendimiento. Sin embargo, su implementación en hardware embebido requerirá optimización adicional o el uso de aceleración especializada.

En síntesis, los objetivos de investigación fueron alcanzados en el marco de una validación técnica de prototipo en entornos controlados. Se diseñó e implementó un sistema funcional que integra visión computacional y retroalimentación auditiva, y se evaluó cuantitativamente su desempeño en términos de precisión, velocidad y costo computacional, identificando con claridad sus fortalezas y limitaciones técnicas.

No obstante, es necesario reconocer que la ausencia de usuarios finales con discapacidad visual en la evaluación constituye una limitación metodológica relevante. Los resultados obtenidos caracterizan el desempeño computacional del sistema, pero no permiten inferir de manera directa la experiencia real de navegación de la población objetivo, dado que la interacción con el entorno y el procesamiento de señales auditivas difieren entre usuarios con y sin discapacidad visual.

En consecuencia, se proponen como líneas futuras la validación del sistema con usuarios reales con discapacidad visual bajo protocolos éticos aprobados, la evaluación en entornos abiertos y dinámicos, la incorporación de nuevas clases al dataset para mejorar la cobertura del sistema y la optimización del hardware hacia plataformas embebidas portátiles.

7 Recomendaciones

Se recomienda explorar alternativas a la síntesis de voz basada en gTTS, dado que este método introduce pausas inevitables en el procesamiento, ya que cada instrucción hablada obliga al sistema a suspender momentáneamente la detección, afectando la fluidez del ciclo percepción-acción. Para mitigar esta limitación, podría implementarse un sistema de retroalimentación auditiva basado en señales acústicas simples, como tonos, patrones rítmicos o variaciones de frecuencia, que comuniquen la presencia y proximidad de obstáculos sin requerir frases completas. Este enfoque ha sido ampliamente utilizado en ayudas electrónicas para personas con discapacidad visual, ya que reduce la latencia y evita la saturación del canal auditivo.

Como alternativa intermedia, pueden evaluarse motores de síntesis de voz locales de baja latencia, como Coqui TTS o Piper, los cuales eliminan la dependencia de servicios externos y reducen los tiempos de respuesta. Sin embargo, incluso estas soluciones presentan retardos superiores en comparación con esquemas de señalización no verbal, por lo que la elección del enfoque debe equilibrar claridad semántica y eficiencia temporal.

En cuanto al procesamiento visual, se recomienda fortalecer el modelo mediante la ampliación y diversificación del conjunto de datos. Incorporar mayor variabilidad en condiciones de iluminación, ángulos de captura, oclusiones parciales y elementos distractores permitirá mejorar la capacidad de generalización del sistema y reducir la aparición de falsos negativos y falsos positivos en escenarios más dinámicos. Asimismo, podría evaluarse el ajuste fino del modelo (fine-tuning) con imágenes capturadas directamente en los entornos de uso previstos.

Respecto al módulo de estereovisión, se sugiere emplear cámaras con mayor línea base o sensores con mejor desempeño en condiciones de bajo contraste, con el fin de obtener mapas de disparidad más estables y precisos. De igual forma, la utilización de interfaces de mayor ancho de banda, como USB 3.0 en lugar de USB 2.0, contribuiría a disminuir la latencia y mejorar la transmisión simultánea de flujos estéreo de alta resolución.

Finalmente, como línea prioritaria de investigación futura, se recomienda llevar a cabo la validación del sistema con usuarios reales con discapacidad visual, bajo protocolos éticos aprobados por un comité institucional. Esta evaluación permitirá analizar aspectos que no pueden ser completamente observados en condiciones controladas sin la participación de usuarios finales, como la interpretación de señales auditivas, la toma de decisiones en movilidad y la experiencia de navegación del usuario. Adicionalmente, se recomienda extender las pruebas a

entornos abiertos y dinámicos, incorporando variaciones en la velocidad de desplazamiento y la densidad de obstáculos, con el fin de determinar la escalabilidad del prototipo hacia aplicaciones prácticas de mayor alcance.

8 Trabajos futuros

Como trabajo futuro prioritario, se propone la validación del sistema con personas con discapacidad visual bajo protocolos éticos aprobados por un comité institucional. Esta etapa permitirá evaluar dimensiones que no pueden ser completamente caracterizadas mediante pruebas técnicas en entornos controlados, como la interpretación real de las instrucciones auditivas, la confianza del usuario en el sistema y su impacto en la autonomía de movilidad. Los resultados de esta validación permitirán definir los ajustes necesarios para avanzar hacia una versión del sistema orientada a contextos de uso real.

Se recomienda profundizar en la optimización del sistema, particularmente en la configuración del módulo estéreo, el ajuste de los parámetros de estereovisión y el perfeccionamiento del modelo de detección, con el objetivo de incrementar la precisión y estabilidad en distintos entornos. Asimismo, la ampliación y diversificación del conjunto de datos de entrenamiento es fundamental para mejorar la capacidad de generalización ante variaciones de iluminación, oclusiones y cambios estructurales del entorno.

De igual forma, es conveniente validar el prototipo en espacios más amplios y no controlados, incorporando escenarios con geometría compleja, mayor densidad de obstáculos y trayectorias menos predecibles. Estas pruebas permitirían evaluar la adaptabilidad del sistema en condiciones más cercanas al uso real y determinar su robustez fuera del contexto experimental.

En relación con la retroalimentación auditiva, se sugiere explorar esquemas basados en patrones sonoros, tonos direccionales o audio espacial, que reduzcan la latencia asociada a la síntesis de voz y proporcionen información más inmediata e intuitiva. Asimismo, la evaluación de motores de síntesis de voz locales de baja latencia podría contribuir a mejorar la fluidez del ciclo percepción–acción.

Desde el punto de vista computacional, se recomienda analizar el desempeño del sistema en distintas configuraciones de hardware, con el fin de identificar plataformas más portátiles y energéticamente eficientes. Esto incluye la evaluación de cámaras estéreo con mayor calidad óptica o mayor ancho de banda, así como dispositivos capaces de ejecutar modelos de visión en tiempo real con menor consumo y mejor gestión térmica.

Finalmente, la incorporación de sensores complementarios, como ultrasonido o LiDAR de corto alcance, podría fortalecer la estimación de profundidad en escenarios donde la visión

por computador presenta limitaciones. De igual manera, el rediseño del prototipo físico, mediante soportes más robustos e integración estructural optimizada del cableado y los módulos electrónicos, contribuiría a mejorar su ergonomía, durabilidad y viabilidad para uso prolongado.

Referencias

- [1] World Health Organization, “World report on vision,” Geneva, 2019. [Online]. Available: <https://www.who.int/publications/i/item/9789241516570>
- [2] S. M. Carrillo-Sierra *et al.*, “Perceptions of Barriers to Inclusion in Students with Disabilities in Higher Education Institutions,” *Societies*, vol. 15, no. 2, Feb. 2025, doi: 10.3390/soc15020037.
- [3] P. P. Kasture, A. Tangade, A. Pole, A. Kumkar, and Y. Jagtap, “Real-Time Object Detection Using Yolo Algorithm for Blind People,” *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 5, no. 2, pp. 152–156, 2021, doi: 10.48175/ijarsct-1191.
- [4] T. Sugashini and G. Balakrishnan, “Visually impaired object segmentation and detection using hybrid Canny edge detector, Hough transform, and improved momentum search in YOLOv7,” *Signal, Image Video Process.*, vol. 18, no. Suppl 1, pp. 251–265, 2024, doi: 10.1007/s11760-024-03149-6.
- [5] A. A. Golroudbari and M. H. Sabour, “Recent Advancements in Deep Learning Applications and Methods for Autonomous Navigation: A Comprehensive Review,” pp. 1–81, 2023, [Online]. Available: <http://arxiv.org/abs/2302.11089>
- [6] Y. Ji, J. Hwang, and E. Y. Kim, “An Intelligent Wheelchair Using Situation Awareness and Obstacle Detection,” *Procedia - Soc. Behav. Sci.*, vol. 97, pp. 620–628, 2013, doi: 10.1016/j.sbspro.2013.10.281.
- [7] K. Bourrier, “A Smart Walker for People with Both Visual and Mobility Impairment,” *A Cult. Hist. Disabil. Long Ninet. Century*, pp. 1–16, 2020, doi: 10.5040/9781350029101.ch-003.
- [8] H. Fernandes, P. Costa, V. Filipe, L. Hadjileontiadis, and J. Barroso, “Stereo vision in blind navigation assistance,” *2010 World Autom. Congr. WAC 2010*, no. October, 2010.
- [9] P. A. Yerlekar, “Artificial Vision for Visually Impaired using YOLO Detection,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 9, no. 5, pp. 1119–1124, 2021, doi: 10.22214/ijraset.2021.34470.
- [10] S. Pokhrel, “IMAGE TO SPEECH CONVERSION FOR VISUALLY IMPAIRED PEOPLE,” *Ayan*, vol. 15, no. 1, pp. 37–48, 2024.
- [11] A. R. Jambhulkar, A. R. Gajera, C. M. Bhavsar, and S. Vatkar, “Real-Time Object Detection and Audio Feedback for the Visually Impaired,” *2023 3rd Asian Conf. Innov.*

- Technol. ASIANCON 2023*, pp. 1–5, 2023, doi: 10.1109/ASIANCON58793.2023.10269899.
- [12] K. Guravaiah, Y. S. Bhavadeesh, P. Shwejan, A. H. Vardhan, and S. Lavanya, “Third Eye: Object Recognition and Speech Generation for Visually Impaired,” *Procedia Comput. Sci.*, vol. 218, pp. 1144–1155, 2022, doi: 10.1016/j.procs.2023.01.093.
- [13] L. E. Sucar and G. Gomez, “Visión Computacional Giovani Gómez Helmholtz Zentrum Munchen,” no. February, p. 186, 2015.
- [14] *Learning Image Processing with OpenCV*. 2015. [Online]. Available: <https://www.packtpub.com/application-development/learning-image-processing-opencv>
- [15] V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning,” pp. 1–31, 2018, [Online]. Available: <http://arxiv.org/abs/1603.07285>
- [16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.
- [17] M. Yaseen, “What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector,” vol. 8, pp. 1–10, 2024, [Online]. Available: <http://arxiv.org/abs/2409.07813>
- [18] Ultralytics, “YOLOv8 - Ultralytics YOLO Docs,” Ultralytics Documentation. Accessed: Apr. 09, 2026. [Online]. Available: <https://docs.ultralytics.com/models/yolov8/>
- [19] Ultralytics, “YOLOv8.” Accessed: Sep. 22, 2024. [Online]. Available: <https://docs.ultralytics.com/#how-can-i-get-started-with-yolo-installation-and-setup>
- [20] R. A. Hamzah and H. Ibrahim, “Literature survey on stereo vision disparity map algorithms,” *J. Sensors*, vol. 2016, 2016, doi: 10.1155/2016/8742920.
- [21] M. E. Cabrera Salceda, “Discapacidad visual,” *Orienta Integr. e inclusión Educ.*, vol. 8, no. 8, p. 15, 2011, [Online]. Available: http://www.sld.cu/galerias/pdf/sitios/rehabilitacion-arte/discapacidad_visual.pdf
- [22] R. de J. Villar Valdés and M. E. Macias Llanes, “Salud visual y sociedad: basamento para los cambios en la formación del especialista en Oftalmología,” *Humanidad. med*, vol. 7, no. 1, pp. 0–0, 2007.
- [23] skylife, “Sistemas de localización en interiores basados en redes inalámbricas.” Accessed: Nov. 06, 2024. [Online]. Available: <https://skylife-eng.com/actualidad/posicionamiento->

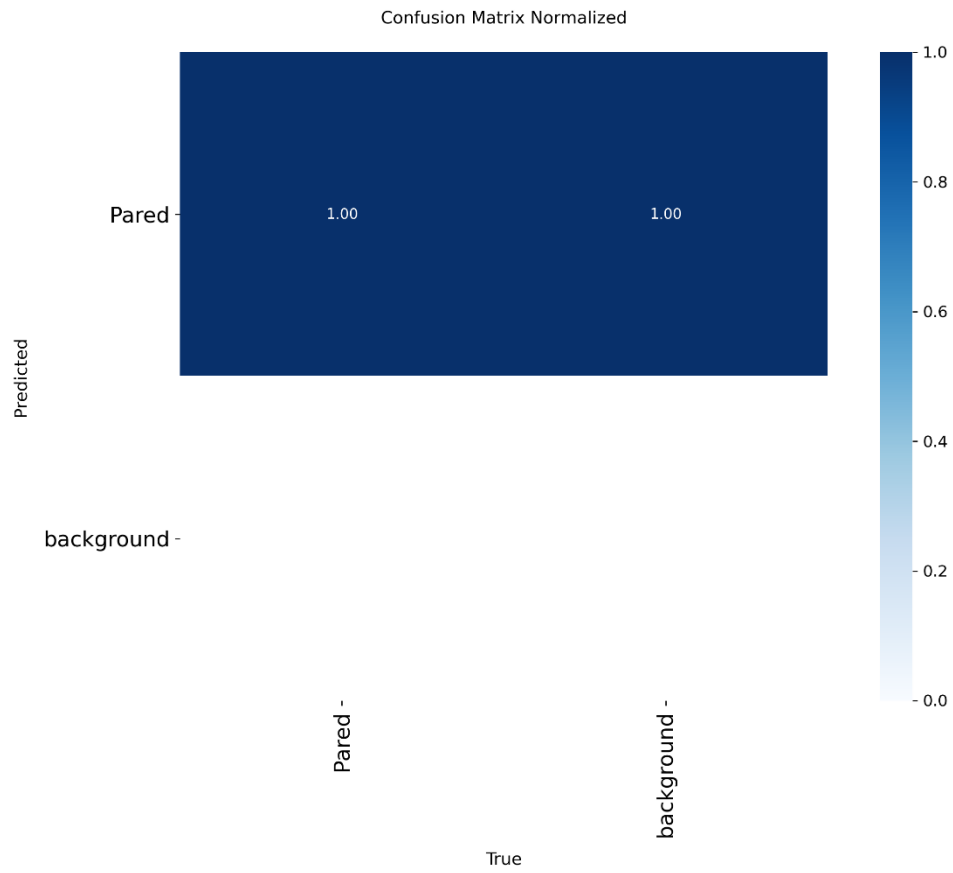
indoor/comparativa-entre-sistemas-de-localización-en-interiores

- [24] “Sistemas de procesamiento de datos”.
- [25] BMJ, “Documento de Posición Oficial,” *Ecoinfo*, pp. 1–4, 2011, [Online]. Available: <https://bit.ly/3bZouEe>
- [26] B. L. Spivey, “Handy Handouts,” *Super Duper Publ.*, no. 374, pp. 2–3, 2012.
- [27] G. Clound, “Conceptos básicos de Cloud Text-to-Speech.” Accessed: Nov. 06, 2024. [Online]. Available: <https://cloud.google.com/text-to-speech/docs/basics?hl=es-419>
- [28] Esri, “Modelos preentrenados de ArcGis,” *Esri*.
- [29] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.
- [30] J. G. Mooney and E. N. Johnson, “A Comparison of Automatic Nap-of-the-earth Guidance Strategies for Helicopters,” *J. F. Robot.*, vol. 23, no. June, pp. 1–17, 2014, doi: 10.1002/rob.
- [31] B. K. Asiedu Asante and H. Imamura, “Towards Robust Obstacle Avoidance for the Visually Impaired Person Using Stereo Cameras,” *Technologies*, vol. 11, no. 6, 2023, doi: 10.3390/technologies11060168.
- [32] A. Zapeta Hernández, G. A. Galindo Rosales, H. J. Juan Santiago, and M. Martínez Lee, “Métricas de rendimiento para evaluar el aprendizaje automático en la clasificación de imágenes petroleras utilizando redes neuronales convolucionales,” *Cienc. Lat. Rev. Científica Multidiscip.*, vol. 6, no. 5, pp. 4624–4637, 2022, doi: 10.37811/cl_rcm.v6i5.3420.
- [33] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.
- [34] M. Hossin and M. N. Sulaiman, “A Review on Evaluation Metrics for Data Classification Evaluations,” *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 01–11, 2015, doi: 10.5121/ijdkp.2015.5201.
- [35] C. J. van Rijsbergen, *Information Retrieval*, 2nd ed. London: Butterworth-Heinemann, 1979.
- [36] R. Borja-Robalino, A. Monleon-Getino, A. Monleón-Getino, and J. Rodellar,

“Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning,” *Rev. Ibérica Sist. e Tecnol. Informação*, vol. 30, no. February, pp. 184–196, 2020, [Online]. Available: <https://www.researchgate.net/publication/342009715>

Anexo B

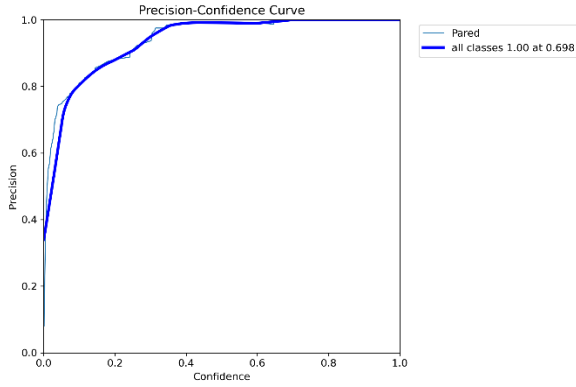
Matriz de confusión normalizada de la instancia estructural (clase Wall)



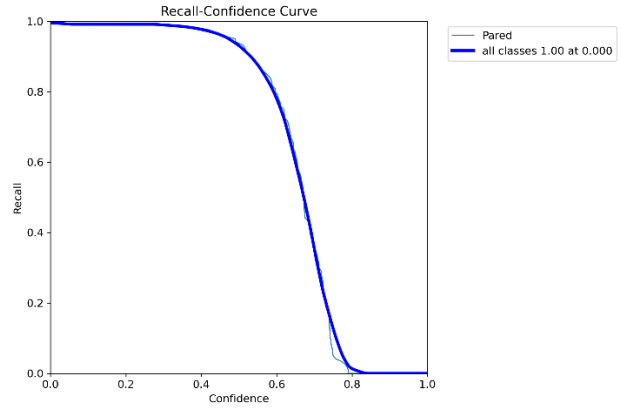
Anexo C

Métricas de validación de la instancia estructural (clase Wall)

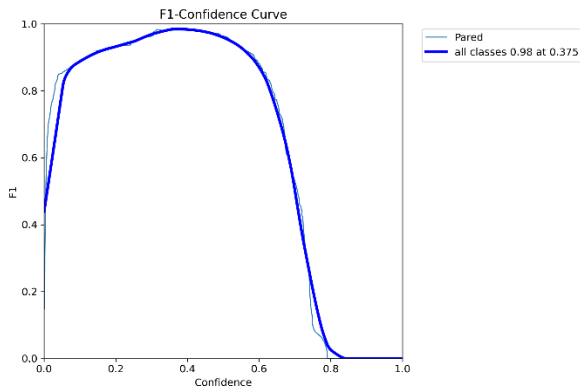
a)



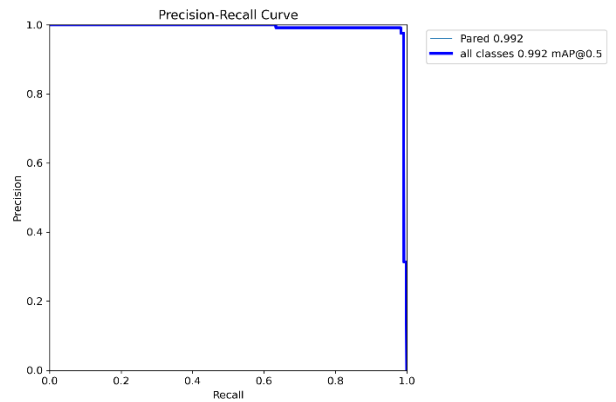
b)



c)



d)



Nota. Curvas de rendimiento para la detección de límites estructurales: (a) Precision-Confidence; (b) Recall-Confidence; (c) F1-Confidence; (d) Precision-Recall. La obtención de un mAP de 0.992 y la estabilidad de las curvas de confianza ratifican la alta especialización del modelo modular para la identificación de paredes, asegurando una base sólida para los algoritmos de navegación.

Anexo D

Resultados Detallados – Escenario 2 (Laboratorio de Física)

Anexo D 1

Tabla de detección de objetos (Criterio A)

Prueba	Objetos enfrentados	Objetos detectados (TP)	No detectados (FN)	Recall (%)
1	9	9	0	100.00
2	10	10	0	100.00
3	10	9	1	90.00
4	10	10	0	100.00
5	10	10	0	100.00
6	10	10	0	100.00
7	7	6	1	85.71
8	9	9	0	100.00
9	10	10	0	100.00
10	10	10	0	100.00
11	10	9	1	90.00
12	10	10	0	100.00
13	10	10	0	100.00
14	10	10	0	100.00
15	10	10	0	100.00
16	10	10	0	100.00
17	9	9	0	100.00
18	10	9	1	90.00
19	10	10	0	100.00
20	10	10	0	100.00

TOTAL	194	190	4	97.93
--------------	------------	------------	----------	--------------

Anexo D 2

Desempeño de clasificación de objetos (Criterio B)

Prueba	Objetos detectados (n)	Clasificados correctamente	Clasificados incorrectamente	Accuracy (%)
		(TP)	(FP)	
1	9	7	2	77.78
2	10	8	2	80.00
3	9	8	1	88.89
4	10	9	1	90.00
5	10	9	1	90.00
6	10	8	2	80.00
7	6	4	2	66.67
8	9	9	0	100.00
9	10	9	1	90.00
10	10	8	2	80.00
11	9	8	1	88.89
12	10	10	0	100.00
13	10	9	1	90.00
14	10	9	1	90.00
15	10	9	1	90.00
16	10	7	3	70.00
17	9	8	1	88.89
18	9	7	2	77.78
19	10	9	1	90.00
20	10	10	0	100.00
TOTAL	190	166	24	87.37

Anexo D 3

Precisión de la estimación de distancias (Estereovisión) sobre objetos detectados (Criterio C)

Prueba	Objetos detectados (n)	Estimados	Estimados	Accuracy (%)
		correctamente (TP)	incorrectamente (FP)	
1	9	9	0	100.00
2	10	8	2	80.00
3	9	9	0	100.00
4	10	10	0	100.00
5	10	8	2	80.00
6	10	9	1	90.00
7	6	6	0	100.00
8	9	9	0	100.00
9	10	10	0	100.00
10	10	9	1	90.00
11	9	7	2	77.78
12	10	10	0	100.00
13	10	7	3	70.00
14	10	6	4	60.00
15	10	9	1	90.00
16	10	9	1	90.00
17	9	8	1	88.89
18	9	7	2	77.78
19	10	9	1	90.00
20	10	8	2	80.00
TOTAL	190	164	26	86.31

Anexo D 4

Resultados de la validación de asignación de zonas espaciales (Criterio D)

Zona	Total	Correctas	Accuracy (%)
9	1	1	100.00
10	3	3	100.00
11	11	11	100.00
12	9	9	100.00
1	7	7	100.00
2	6	6	100.00
3	3	3	100.00
Total	40	40	100.00

Anexo D 5

Resultados de la evaluación del sistema de navegación por visión computacional (YOLO + estereovisión)

Prueba	Objetos enfrentados	Aciertos (TP)	Fallos Parciales (EP)	No detectados (FN)
1	9	7	2	0
2	10	6	4	0
3	10	8	1	1
4	10	9	1	0
5	10	7	3	0
6	10	7	3	0
7	7	4	2	1
8	9	9	0	0
9	10	9	1	0
10	10	7	3	0
11	10	6	3	1

12	10	10	0	0
13	10	6	4	0
14	10	5	5	0
15	10	8	2	0
16	10	6	4	0
17	9	7	2	0
18	10	7	3	0
19	10	8	2	0
20	10	8	2	0
TOTAL	194	147	44	3

Anexo E

Resultados Detallados – Escenario 3 (Sexto piso Edificio San Francisco)

Anexo E 1

Tabla de detección de objetos (Criterio A)

Prueba	Objetos enfrentados	Objetos detectados (TP)	No detectados (FN)	Recall (%)
1	13	13	0	100.00
2	6	5	1	83.33
3	13	13	0	100.00
4	10	10	0	100.00
5	12	12	0	100.00
6	12	12	0	100.00
7	7	6	1	85.71
8	12	12	0	100.00
9	10	9	1	90.00
10	13	13	0	100.00
11	13	13	0	100.00
12	12	12	0	100.00

13	7	7	0	100.00
14	11	11	0	100.00
15	11	11	0	100.00
16	11	11	0	100.00
17	13	13	0	100.00
18	12	12	0	100.00
19	13	13	0	100.00
20	13	13	0	100.00
TOTAL	224	221	3	98.66

Anexo E 2

Desempeño de clasificación de objetos (Criterio B)

Prueba	Objetos detectados (n)	Clasificados	Clasificados	Accuracy (%)
		correctamente (TP)	incorrectamente (FP)	
1	13	11	2	84.62
2	5	5	0	100.00
3	13	12	1	92.31
4	10	9	1	90.00
5	12	10	2	83.33
6	12	12	0	100.00
7	6	6	0	100.00
8	12	11	1	91.67
9	9	9	0	100.00
10	13	12	1	92.31
11	13	12	1	92.31
12	12	12	0	100.00
13	7	7	0	100.00
14	11	11	0	100.00
15	11	10	1	90.91

16	11	11	0	100.00
17	13	13	0	100.00
18	12	12	0	100.00
19	13	11	2	84.62
20	13	12	1	92.31
TOTAL	221	208	13	94.12%

Anexo E 3

Precisión de la estimación de distancias (Estereovisión) sobre objetos detectados (Criterio C)

Prueba	Objetos detectados (n)	Estimados	Estimados	Accuracy (%)
		correctamente (TP)	incorrectamente (FP)	
1	13	11	2	84.61
2	5	5	0	100.00
3	13	10	3	76.92
4	10	8	2	80.00
5	12	11	1	91.67
6	12	12	0	100.00
7	6	6	0	100.00
8	12	11	1	91.67
9	9	7	2	77.78
10	13	12	1	92.31
11	13	12	1	92.31
12	12	12	0	100.00
13	7	6	1	85.71
14	11	11	0	100.00
15	11	8	3	72.73
16	11	10	1	90.91
17	13	11	2	84.61
18	12	12	0	100.00

19	13	11	2	84.61
20	13	10	3	76.92
TOTAL	221	196	25	88.68 %

Anexo E 4

Resultados de la validación de asignación de zonas espaciales (Criterio D)

Zona	Total	Correctas	Accuracy (%)
9	0	0	N/A
10	4	4	100.00
11	3	3	100.00
12	12	12	100.00
1	11	11	100.00
2	9	9	100.00
3	1	1	100.00
Total	40	40	100.00

Anexo E 5

Resultados de la evaluación del sistema de navegación por visión computacional (YOLO + estereovisión)

Prueba	Objetos	Aciertos	Errores	No Detectados
	Enfrentados	Totales (TP)	Parciales (EP)	(FN)
1	13	11	2	0
2	6	5	0	1
3	13	10	3	0
4	10	8	2	0
5	12	10	2	0
6	12	12	0	0
7	7	6	0	1

8	12	11	1	0
9	10	7	2	1
10	13	12	1	0
11	13	12	1	0
12	12	12	0	0
13	7	6	1	0
14	11	11	0	0
15	11	8	3	0
16	11	10	1	0
17	13	11	2	0
18	12	12	0	0
19	13	11	2	0
20	13	10	3	0
TOTAL	224	195	26	3

 <p>UNIVERSIDAD CESMAG NIT: 800.109.387-7 VIA AEREA INTERNACIONAL</p>	CARTA DE ENTREGA TRABAJO DE GRADO O TRABAJO DE APLICACIÓN – ASESOR(A)	CÓDIGO: AAC-BL-FR-032
		VERSIÓN: 1
		FECHA: 09/JUN/2022

San Juan de Pasto, 29 de abril de 2026

Biblioteca
REMIGIO FIORE FORTEZZA OFM. CAP.
Universidad CESMAG
Pasto


Saludo de paz y bien.

Por medio de la presente se hace entrega del Trabajo de Grado / Trabajo de Aplicación denominado "Sistema de navegación en interiores asistido por visión computacional orientado a personas con discapacidad visual" presentado por el (los) autor(es) Manuel Stiven Leon Males y Luis Felipe Paz Caicedo del Programa Académico Ingeniería Electrónica al correo electrónico biblioteca.trabajosdegrado@unicesmag.edu.co. Manifiesto como asesor(a), que su contenido, resumen, anexos y formato PDF cumple con las especificaciones de calidad, guía de presentación de Trabajos de Grado o de Aplicación, establecidos por la Universidad CESMAG, por lo tanto, se solicita el paz y salvo respectivo.

Atentamente,



José Camilo Eraso Guerrero
1085290125
Ingeniería Electrónica
3187930965
jceraso@unicesmag.edu.co

 UNIVERSIDAD CESMAG <small>NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</small>	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

INFORMACIÓN DEL (LOS) AUTOR(ES)	
Nombres y apellidos del autor: Manuel Stiven Leon Males	Documento de identidad: 1004696684
Correo electrónico: msleon.6684@unicesmag.edu.co	Número de contacto: 3186453057
Nombres y apellidos del autor: Luis Felipe Paz Caicedo	Documento de identidad: 1193082193
Correo electrónico: lfpaz.2193@unicesmag.edu.co	Número de contacto: 3132391633
Nombres y apellidos del autor:	Documento de identidad:
Correo electrónico:	Número de contacto:
Nombres y apellidos del asesor: José Camilo Eraso Guerrero	Documento de identidad: 1085290125
Correo electrónico: jceraso@unicesmag.edu.co	Número de contacto: 3187930965
Título del trabajo de grado: Sistema de navegación en interiores asistido por visión computacional orientado a personas con discapacidad visual	
Facultad y Programa Académico: Facultad de ingeniería / Ingeniería Electrónica	

En mi (nuestra) calidad de autor(es) y/o titular (es) del derecho de autor del Trabajo de Grado o de Aplicación señalado en el encabezado, confiero (conferimos) a la Universidad CESMAG una licencia no exclusiva, limitada y gratuita, para la inclusión del trabajo de grado en el repositorio institucional. Por consiguiente, el alcance de la licencia que se otorga a través del presente documento, abarca las siguientes características:

- a) La autorización se otorga desde la fecha de suscripción del presente documento y durante todo el término en el que el (los) firmante(s) del presente documento conserve (mos) la titularidad de los derechos patrimoniales de autor. En el evento en el que deje (mos) de tener la titularidad de los derechos patrimoniales sobre el Trabajo de Grado o de Aplicación, me (nos) comprometo (comprometemos) a informar de manera inmediata sobre dicha situación a la Universidad CESMAG. Por consiguiente, hasta que no exista comunicación escrita de mi(nuestra) parte informando sobre dicha situación, la Universidad CESMAG se encontrará debidamente habilitada para continuar con la publicación del Trabajo de Grado o de Aplicación dentro del repositorio institucional. Conozco(conocemos) que esta autorización podrá revocarse en cualquier momento, siempre y cuando se eleve la solicitud por escrito para dicho fin ante la Universidad CESMAG. En estos eventos, la Universidad CESMAG cuenta con el plazo de un mes después de recibida la petición, para desmarcar la visualización del Trabajo de Grado o de Aplicación del repositorio institucional.
- b) Se autoriza a la Universidad CESMAG para publicar el Trabajo de Grado o de Aplicación en formato digital y teniendo en cuenta que uno de los medios de publicación del repositorio institucional es el internet, acepto(amos) que el Trabajo de Grado o de Aplicación circulará con un alcance mundial.

 UNIVERSIDAD CESMAG <small>RIT. 000 109.387-7 VICEPRESIDENCIA</small>	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

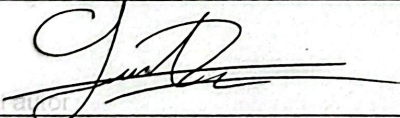
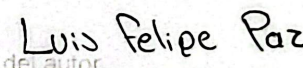
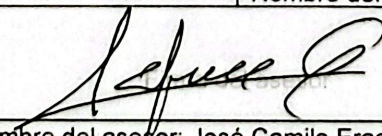
- c) Acepto (aceptamos) que la autorización que se otorga a través del presente documento se realiza a título gratuito, por lo tanto, renuncio(amos) a recibir emolumento alguno por la publicación, distribución, comunicación pública y/o cualquier otro uso que se haga en los términos de la presente autorización y de la licencia o programa a través del cual sea publicado el Trabajo de grado o de Aplicación.
- d) Manifiesto (manifestamos) que el Trabajo de Grado o de Aplicación es original realizado sin violar o usurpar derechos de autor de terceros y que ostento(amos) los derechos patrimoniales de autor sobre la misma. Por consiguiente, asumo(asumimos) toda la responsabilidad sobre su contenido ante la Universidad CESMAG y frente a terceros, manteniéndose indemne de cualquier reclamación que surja en virtud de la misma. En todo caso, la Universidad CESMAG se compromete a indicar siempre la autoría del escrito incluyendo nombre de(los) autor(es) y la fecha de publicación.
- e) Autorizo(autorizamos) a la Universidad CESMAG para incluir el Trabajo de Grado o de Aplicación en los índices y buscadores que se estimen necesarios para promover su difusión. Así mismo autorizo (autorizamos) a la Universidad CESMAG para que pueda convertir el documento a cualquier medio o formato para propósitos de preservación digital.

NOTA: En los eventos en los que el trabajo de grado o de aplicación haya sido trabajado con el apoyo o patrocinio de una agencia, organización o cualquier otra entidad diferente a la Universidad CESMAG. Como autor(es) garantizo(amos) que he(hemos) cumplido con los derechos y obligaciones asumidos con dicha entidad y como consecuencia de ello dejo(dejamos) constancia que la autorización que se concede a través del presente escrito no interfiere ni transgrede derechos de terceros.

Como consecuencia de lo anterior, autorizo(autorizamos) la publicación, difusión, consulta y uso del Trabajo de Grado o de Aplicación por parte de la Universidad CESMAG y sus usuarios así:

- Permiso(permitimos) que mi(nuestro) Trabajo de Grado o de Aplicación haga parte del catálogo de colección del repositorio digital de la Universidad CESMAG por lo tanto, su contenido será de acceso abierto donde podrá ser consultado, descargado y compartido con otras personas, siempre que se reconozca su autoría o reconocimiento con fines no comerciales.

En señal de conformidad, se suscribe este documento en San Juan de Pasto a los 29 días del mes de abril del año 2026

	
<small>Firma del autor</small>	<small>Firma del autor</small>
Nombre del autor: Manuel Stiven Leon Males	Nombre del autor: Luis Felipe Paz Caicedo
<small>Firma del autor</small>	<small>Firma del autor</small>
Nombre del autor:	Nombre del autor:
	
Nombre del asesor: José Camilo Eraso Guerrero	