PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE INGENIERÍA

ESCUELA DE SISTEMAS Y COMPUTACIÓN



TRABAJO PREVIO A LA OBTENCIÓN DEL TITULO DE

INGENIERO EN SISTEMAS Y COMPUTACIÓN

DISEÑO E IMPLEMENTACIÓN DE UN PROTOTIPO BASADO EN INTELIGENCIA ARTIFICIAL PARA EL RECONOCIMIENTO DE OBJETOS EN EL ENTORNO DE UNA PERSONA NO VIDENTE

AUTOR

BENALCÁZAR ACOSTA CÉSAR ANDRÉS

QUITO DM, NOVIEMBRE 2022

Dedicatoria

A Hazel el regalo más grande que Dios
me supo entregar, mi mayor motivación
para cumplir cada objetivo.
A Maytee quien con su ayuda incondicional
me ha permitido crecer y ser una mejor persona.

Agradecimientos

A mis padres, quienes han sido un
pilar fundamental para mi educación
y un ejemplo a seguir en cada meta propuesta.

A mis hermanas, amigos, docentes y a todas las
personas que de una u otra forma me ayudaron
a culminar mis estudios universitarios.

Contenido

Justificación 9
Antecedentes 10
Planteamiento del Problema 11
Objetivos 13
Objetivo General 13
Objetivos Específicos 13
Capítulo 2: Marco de Referencia 14
2.1 Marco Conceptual Discapacidad Visual 14
2.1.1 Discapacidad Visual 14
2.1.2 Ceguera 14
2.1.3 Disminución Visual 15
2.2 Marco Conceptual Tecnologías de Información 16
2.2.1 Visión Artificial 16
2.2.2 Inteligencia 18
2.2.3 Inteligencia Artificial 18
2.2.4 Redes Neuronales 21
2.2.5 Proceso de Detección de Objetos 24
Algoritmos 24
2.2.6 Hardware 26
2.2.7 Software 29
2.2.8 Metodología 31
Capítulo 3: Desarrollo del Prototipo 33
3.1 Identificación de objetivos de detección con el uso de software libre de aprendizaje automático y bibliotecas de visión artificial 33
3.1.1 Características principales del prototipo 33
3.1.2 Análisis e interpretación de encuestas 34
3.1.3 Requerimientos del usuario 38
3.1.4 Características Técnicas Obtenidas 39
3.2 Diseño de un prototipo de gafas mediante el uso de una cámara, un dispositivo de audio y el uso de un computador de placa reducida para beneficio de personas no videntes. 41
3.2.1 Proceso de reconocimiento de objetos 42
3.2.2 Entrenamiento del modelo de detección (personalizado) 43
3.3 Implementación del prototipo de visión artificial basado en software libre y hardware de bajo costo 48

3.3.1 Software del sistema 48
3.3.2 Configuración Raspberry PI 49
3.3.3 Batería Portatil 4000mAh (USB-C) 52
3.3.4 Evaluación del prototipo 53
3.3.5 Pruebas en personas con discapacidad visual 55
3.3.6 Costo referencial del prototipo 56
Capítulo 4: Conclusiones y Recomendaciones 57
4.1 Conclusiones 57
4.2 Recomendaciones 58
Bibliografía: 59
Anexos 60

Índice de Tablas

Tabla 1. Especificaciones Raspberry PI 4 B 27
Tabla 2. Pregunta 1 Encuesta Identificación de objetivos y requerimientos del usuario..... 35
Tabla 3. Pregunta 2 Encuesta Identificación de objetivos y requerimientos del usuario..... 35
Tabla 4. Pregunta 3 Encuesta Identificación de objetivos y requerimientos del usuario..... 35
Tabla 5. Pregunta 4 Encuesta Identificación de objetivos y requerimientos del usuario..... 36
Tabla 6. Pregunta 5 Encuesta Identificación de objetivos y requerimientos del usuario..... 36
Tabla 7. Pregunta 6 Encuesta Identificación de objetivos y requerimientos del usuario..... 37
Tabla 8. Pregunta 7 Encuesta Identificación de objetivos y requerimientos del usuario..... 37
Tabla 9. Pregunta 8 Encuesta Identificación de objetivos y requerimientos del usuario..... 38
Tabla 10. Requerimientos del Usuario 39
Tabla 11. Características técnicas del prototipo 40
Tabla 12. Identificación de objetos 40
Tabla 13. Comparación de Arquitectura EfficentDet..... 46
Tabla 14. Evaluación del prototipo en objetos reales..... 54
Tabla 15. Respuesta promedio detección 55
Tabla 16. Resultados pruebas aplicadas a personas 56
Tabla 17. Costo referencial del prototipo 56

Índice de Figuras

Figura 1. Componentes del prototipo de visión artificial 16
Figura 2. Estructura de una Neurona Artificial 22
Figura 3. Estructura de una Red Neuronal Artificial 22
Figura 4. Diagrama general de funcionamiento 33
Figura 5. Esquemático conexión Raspberry PI a los dispositivos externos 41
Figura 6. Diagrama general de flujo 42
Figura 7. Diagrama de flujo correspondiente al algoritmo de reconocimiento de objetos. 43
Figura 8. Set imágenes del objeto manzana 44
Figura 9. Software de anotación de imágenes LabelImg 44
Figura 10. Archivo XML resultante del etiquetado de imágenes 45
Figura 11. Librerías importadas para la ejecución del código 45
Figura 12. Selección de elementos para entrenamiento y validación. 46
Figura 13. Arquitectura seleccionada para el entrenamiento 46
Figura 14. Parámetros establecidos para el entrenamiento 47
Figura 15. Validación de datos 47
Figura 16. Exportar el modelo entrenado 47
Figura 17. Prototipo de gafas de visión artificial. 48
Figura 18. Modificar la memoria en el swapfile 49
Figura 19. Configuración de Raspberry PI 50
Figura 20. Habilitar conexión de la cámara 50
Figura 21. Interfaz de usuario, Botón R 51
Figura 22. Botón encendido/apagado. Botón (S) 51
Figura 23. Periféricos de entrada y salida removibles de la Raspberry Pi. 52
Figura 24. Batería 4000mAh, 5V 2.4A (USB-C) 53
Figura 25. Foto prueba del dispositivo con personas 55

Introducción

El propósito del primer capítulo es el de ofrecer una inducción al presente trabajo para obtención de la titulación. Durante su desarrollo se describen: Justificación, la cual expone las razones que motivaron la realización de esta investigación, explicando la importancia y los motivos para realizar el trabajo. Los antecedentes, que son un compendio de saberes, análisis, estudios y proyectos previos, sean a nivel nacional o internacional. El planteamiento del problema, identifica la situación del tema a ser resuelto. Y por último; objetivo general y objetivos específicos, otorgan el enfoque y las metas a cumplirse a través de la elaboración de los mismos para obtener los resultados deseados.

Justificación

Conforme a datos proporcionados por el Consejo Nacional para la Igualdad de Discapacidades (CONADIS); en Ecuador existen alrededor de 54377 personas que padecen de discapacidad visual, cifra que equivale al 11,54% del total de la población registrada con discapacidad a nivel nacional (CONADIS, 2022) [1]. Debido al alto porcentaje de población invidente el proyecto adquiere relevancia fundamentándose en la necesidad de brindar soluciones, a través de la tecnología aplicada en beneficio de las personas con discapacidad visual.

Uno de las dificultades más frecuentes con las que deben batallar las personas con deficiencia visual es el hecho de no poder desplazarse de un lugar a otro de manera independiente, puesto que, tanto la orientación como la movilidad, son factores limitantes que se encuentran íntimamente relacionados.

En la actualidad, el constante crecimiento de las tecnologías basadas en hardware y software facilita la obtención, transferencia, procesamiento y archivo de datos de manera ágil, de esta forma la inteligencia artificial pasa a ser una de las más significativas para la detección de objetos y de envío de alertas al aparato auditivo. Siendo capaces de reemplazar el uso de bastones, mascotas que sirven de guía y otras herramientas utilizadas por personas no videntes para su habitual desenvolvimiento.

1 CONADIS: Consejo Nacional para la Igualdad de las Discapacidades

Mediante el presente trabajo investigativo se tiene la expectativa de alcanzar una propicia aceptación del proyecto, por parte de los usuarios. Así como el cumplimiento de los objetivos trazados, se aspira obtener resultados positivos sin escatimar la posibilidad de que este trabajo sirva de base para futuros estudios de carácter científico en beneficio de las personas más vulnerable.

Antecedentes

Para contextualizar el tema, se debe tener conocimiento de que la visualización es una habilidad que adquieren las personas invidentes para construir mentalmente una idea no tan concreta de su entorno, se por la utilización de representaciones verbales, búsquedas para obtener información mediante funciones táctiles y de la kinésica[2]. En resumen: La o las imágenes que se forman dentro de la mente del no vidente, en cuanto a su orientación, se conoce como: visualización. Definiendo como orientación a la noción definida de la ubicación de una persona paralela a los objetos que estén situados a su alrededor. Para lo cual se deben tomar en cuenta lo siguiente:

- La persona debe tener conocimiento previo de las circunstancias o características que le rodean, entre estos están; los objetos que hay en el sitio, además de su tamaño, estructura y ubicación de los mismos, en resumen, se refiere a la búsqueda del conocimiento del ambiente.
- La distribución de los objetos por si la persona con deficiencia visual cambia de una ubicación a otra.
- Tener en consideración que, la movilidad es el acto de trasladarse dentro de un espacio.

Explicado de otra forma; es la destreza que adquieren las personas invidentes para cambiar de ubicación dentro de su hábitat.

Los individuos con insuficiencia visual captan información de su entorno por medio de los sentidos. Al carecer del sentido de la visión, tienden a desarrollar sus demás sentidos como son: oído, olfato, tacto y gusto. Vale acotar que existen sentidos adicionales como el kinestésico.

Dentro de los proyectos creados con antelación al actual, se hallan contemplados los siguientes:

Uno de los aportes tecnológicos realizados recientemente está la invención de gafas inteligentes diseñadas en la Universidad de Oxford, las cuales sacan provecho de la visión excedente de los no videntes para que les sea posible su orientación mediante entornos no explorados por ellos. Dichas gafas están compuestas por un sistema de cámaras y un software que tienen la función de detectar objetos cercanos para que el usuario logre reconocerlos.

La creación de este prototipo se le atribuye a Stephen Hicks, quien diseñó este modelo, el cual se halla en fase de financiación para obtener una producción a gran escala. Si se consigue el tan anhelado financiamiento, este mismo año, podrían empezar a comercializarse. Su precio sería equivalente al de un celular común.

En Reino Unido, varios alumnos de la Universidad de Birmingham desarrollaron un bastón con tecnología XploR y un GPS[3]. Por su parte XploR, tiene la función de reconocimiento facial, es decir que detecta las caras de las personas en una distancia de nueve a diez metros, aproximadamente. Cuando una persona conocida es identificada se activa a través de la cámara. Mientras que el GPS sirve de guía para el usuario se guíe hacia la persona, cuyo rostro fue detectado, a través de un sistema de audio. Adigbo Steve, uno de los creadores de este bastón identificador, se basa en su entorno personal y manifestó: “Mi abuelo es ciego y sé lo útil que este dispositivo podría ser para él".

Por su parte, en la Universidad Miguel Hernández (UMH) de España, integrantes del equipo de investigación desarrollaron un innovador bastón electrónico para prestar ayudar en el desplazamiento de personas que carecen de visión normal. El cual se fundamenta en un sistema que formula un aviso de vibración ubicado en la muñeca que anuncia la detección de objetos que pudieran ser un peligro inminente para la persona invidente. Dicha detección de objetos se lleva a cabo a través de un componente de sensores adaptados a un bastón común y corriente.

También se creó un pequeño y ligero aparato electrónico que aporta en la movilidad, debido a su función sensorial es capaz de emitir señales sonoras y táctiles, denominado Ray, el cual fue desarrollado como complemento del bastón largo. Funciona a manera de un ultrasonido, emitiendo una especie de advertencia (como los rayos de luz que salen de una linterna) la misma que permite detectar obstáculos de manera más rápida y logra que la persona en cuestión pueda hacerse una mejor idea de su entorno. Ray es muy práctico, cabe en cualquier bolsillo.

---

[2] Kinesica: es una disciplina independiente de la psicología dedicada al estudio del lenguaje de la comunicación no verbal

Funciona con apenas dos baterías AAA de larga duración. Es capaz de reconocer obstáculos de hasta 2,85m de distancia. La existencia de dichos obstáculos se da a conocer mediante la emanación de una señal auditiva o puede escogerse el modo vibrador, esto queda a libre elección del usuario.

Google Glass, incorporó un sistema, Lazzus[4], que se encarga de anunciar detalles del trayecto por donde transite el invidente en un determinado momento , así como; letreros de locales comerciales, la ubicación de señaléticas o cualquier novedad acerca del tráfico. Lo cual facilita llegar al destino deseado.

[3] GPS: Sistema de posicionamiento global
[4] Lazzus: es un asistente que acompaña a las personas ciegas y deficientes visuales en sus desplazamientos.

Lazzus estará disponible para dispositivos con sistema iOS y Android siempre y cuando la ONCE otorgue su validación. Este diseño utiliza el GPS para reconocer dónde está situado el usuario y hace uso de la brújula para establecer la orientación y su campo de visión. Se obtiene resultados, dentro de un rango de veinte metros, a más de poder describir lo que le interesa y necesita el invidente por medio de señales sonoras o de mensajes de voz.

El usuario tiene la opción de configurar las instancias que desee que sean descritas, gracias a las bases de datos e informaciones halladas a través de Foursquare, Tom Tom Go, entre otras. Este GPS no solo anuncia verbalmente los nombres de las calles, los cruces y los puntos de interés mientras el usuario camina, sino que además le permite planear y seguir rutas. El dispositivo está compuesto por un asistente digital personal externo (PDA[5]), un receptor GPS y un altavoz, que están acoplados vía bluetooth.

Este aparato permite que las personas con complicaciones de visión logren encontrar de manera fácil las entradas de los establecimientos comerciales. Este dispositivo funciona de la siguiente manera: cuando el invidente se esté desplazando dentro del perímetro de señal del Step Hear, el mecanismo que debe estar está ubicado en alguna parte de la extremidad superior del usuario se activa y vibra o suena, posee un sistema auditivo que hace más fácil su orientación, por consiguiente; luego de que el botón es pulsado se activa el Step Hear comienza a orientar a la persona hacia la entrada de las tiendas, proporcionando información sobre los productos que ofrece cada local. Es un instrumento hecho para que los no videntes puedan desenvolverse de manera independiente en la calle.

Los sistemas de hardwares libres permiten la creación y compartición de diseños de dispositivos electrónicos, y programas como los llamados "drivers", aportan con libertades de uso, compartición, función para modificar fuentes y para distribuirlas.

Un diseño de hardware libre es aquel que ofrece libertades del software tanto en su diseño, como en el PCB[6] a más de estar presente en el archivo de fábrica.

Ventajas más importantes del hardware libre:

- En el plano económico, una de las ventajas para el usuario es que los precios se reducen ya que no posee fines comerciales o de lucro.
- La ventaja para el desarrollador es que se obtiene un acceso directo al código de diseños y librerías, lo cual da paso para personalizar y mejorar la aplicación

[5] PDA: Asistente Digital Personal

[6] PCB: Placa de Circuito Impreso

Mientras que, la Raspberry PI, fue ideada en 2006, sin embargo fue lanzada al mercado en seis años después, puntualmente en febrero de 2012. Desarrollada por un grupo de alumnos y catedráticos de la Universidad de Cambridge. Se asemeja una computadora, posee el tamaño de una tarjeta de crédito, la cual se conecta a un monitor y a un teclado. Se presenta como una placa que abarca varios elementos necesarios en un ordenador común. Proporciona muchas de las herramientas que un PC de realiza, tales como: procesadores de texto, reproductor vídeos en HD, hojas de cálculo e inclusive juegos.

El objetivo es promover la enseñanza de la Informática a niñas y niños. De hecho, en enero de este año Google concedió más de 15.000 Raspberry Pi hacia planteles educativos en Reino Unido.

La placa posee diversos puertos y entradas, dos de USB, uno de Ethernet y uno salida HDMI[7], los cuales permiten conectar el miniordenador a otros dispositivos, teclados, ratones y pantallas. Además, posee un System on Chip que contiene un procesador ARM[8], mismo que marcha a 700 Mhz, también tiene un procesador gráfico VideoCore IV, capacidad de memoria RAM de 512. Además, es posible instalar sistemas operativos libres a través de una tarjeta SD.

[7] HDMI: Interfaz multimedia de alta definición

[8] ARM: Tipo de arquitectura para fabricar microprocesadores.

Planteamiento del Problema

Los problemas visuales intervienen en la vida de los seres humanos, ya sea: física, mental, emocional, educativa y/o profesionalmente. Dicha problemática puede obstaculizar el hecho de que la persona invidente no pueda: adquirir conocimientos, déficit en sus actividades cotidianas, causar un mal flujo en la comunicación, impedir su orientación y movilidad con normalidad, etc. Debido a aquellos inconvenientes las personas invidentes suplen esa carencia utilizando la información que obtienen de sus demás sentidos como son: la audición para orientarse por medio del oído, identificación de objetos mediante el tacto, su sentido olfativo para guiarse por olores, y así sucesivamente.

Se considera como discapacidad visual a la privación, disminución o desperfectos en el sentido de la visión. Por los tanto se entiende "discapacidad visual" al mal funcionamiento del sentido de la vista, sea éste parcial o total. De las cuales se extraen las siguientes clases:

- La ausencia de respuesta visual es conocida como "amaurosis" o ceguera total.
- Si se tiene un puntaje de 1/10 en agudeza visual en el ojo de mayor visión, es considerada como "ceguera legal".
- Un 3/10 de agudeza visual simboliza una disminución visual.
- Mientras que, la visión subnormal corresponde a la pérdida del campo visual.

Según datos emitidos por la Organización Mundial de la Salud (OMS), se calcula que mil trecientos millones de personas tienen problemas de índole visual (a nivel mundial), cifra de la que treinta y nueve millones padecen de ceguera y doscientos cuarenta y seis millones poseen deficiencia visual grave, de las cuales, la gran mayoría provienen de países subdesarrollados; con precario nivel de nutrición, inadecuados servicios sanitarios, con escasas políticas inclusivas para no videntes, a más de problemas de educación, agua no potable e insalubridad, son factores que aumentan la cifra de personas con enfermedades, atrofias y anomalías en la visión. (OMS, 2021).

Las personas no videntes no pueden desplazarse con seguridad en la calle ni en sus respectivas viviendas, ya que algunos aparatos que coexisten en el mercado como bastón blanco, las mascotas guía y otros medios no cuentan con un sistema de detección de objetos que están ubicados de la cintura para arriba, lo cual puede ocasionar que el usuario sea víctima de caídas o tropiezos y las probabilidades de que sucedan accidentes con algún objeto aéreo o que esté ubicado a la altura de la parte superior de su cuerpo, van en aumento. La gran mayoría de personas que se encuentran dentro de éstas estadísticas, no cuentan con ninguna clase de

dispositivo que proporcione visión artificial y que puedan servir de ayuda para su orientación y movimiento, las personas afecciones visuales no pueden desplazarse y caminar con seguridad, Razones suficientes para idear la manera de cooperar con el mejoramiento de la calidad de vida de los invidentes, a través del presente trabajo.

Objetivos

Objetivo General

- Diseñar e implementar un prototipo de gafas basadas en inteligencia artificial para el reconocimiento de objetos en el entorno de una persona no vidente.

Objetivos Específicos

- Establecer los objetivos de detección con el uso de software libre de aprendizaje automático (Tensorflow) y bibliotecas de visión artificial (OPENCV)
- Diseñar un prototipo de gafas mediante el uso de una cámara, incorporar un dispositivo de audio y la utilización de un ordenador con placa reducida (Raspberry PI)
- Implementar un prototipo de visión artificial fundamentado en software libre y hardware de bajo costo.

Capítulo 2: Marco de Referencia

El presente capítulo contiene la descripción de la base teórica, que permitirán un mayor entendimiento de las características técnicas y conceptos esenciales que permitirán establecer la metodología de trabajo y las herramientas que permitirán el desarrollo del proyecto.

2.1 Marco Conceptual Discapacidad Visual

2.1.1 Discapacidad Visual

Según la autora María Socorro Quintana, la discapacidad visual no debe ser entendida con únicamente cerrar los ojos o con la simulación inmediata de una visión disminuida, puesto que no ver implica poner en juego estrategias auditivas, gustativas, táctiles y olfativas; si existe un remanente visual, implica apoyarse de manera intersensorial para conocer el entorno; percatarse de formas, colores, sucesos, fenómenos naturales y desenvolverse normalmente en el grupo social, para formar parte del mismo.

La discapacidad visual se determina básicamente en la agudeza visual y el campo visual. Se hace alusión a "discapacidad visual" siempre y cuando exista una rebaja significativa de la agudeza visual aún con el uso de lentes, o a su vez, una disminución significativa del campo visual.

2.1.2 Ceguera

Desde la perspectiva de la Organización Mundial de la Salud (OMS), la ceguera corresponde a la imposibilidad de ver. Es una situación biológica de ciertos individuos, que es manifestada como una precaria o ausente visión. Para una mejor comprensión de su definición se tiene varios conceptos como:

- Agudeza visual: capacidad para distinguir formas, detalles y colores a una distancia específica.
- Campo visual: capacidad para percibir objetos en un ángulo de 180 grados mirando de frente, de izquierda a derecha y aproximadamente 140 grados de arriba hacia abajo, por la forma del ojo.
- Visión cercana: la que permite reconocer detalles finos como los caracteres de lecto-escritura, gestos del rostro y movimientos del cuerpo que posibilitan la comunicación gestual.

Según la enciclopedia médica MedlinePlus la ceguera se refiere a una pérdida de la visión que no puede ser corregida con anteojos comunes o con lentes de contacto. Además, se añade la siguiente división:

- La ceguera parcial, es cuando la persona posee una visión muy limitada.
- La ceguera completa, definida así cuando la persona no puede ver nada, ni siquiera algún

rasgo de luz. No está por demás acotar que la mayoría de las personas que utilizan el término "ceguera" se están expresando sobre ceguera completa.

En MedlinePlus, también se manifiesta que las personas con una visión inferior a 20/200, inclusive si utilizan anteojos convencionales o lentes de contacto, se consideran categóricamente ciegas.

2.1.3 Disminución Visual

La OMS, reconoce como baja visión al campo visual disminuido hasta en 10 grados, o agudeza visual de 20/70 hasta percepción de luz.

Por su parte, para la American Academy of Ophtamology la disminución de la visión es una pérdida parcial de visión transitoria o permanente.

Las posibles causas para que se presente una disminución visual son:

- Accidentes o lesiones producidas en la superficie del ojo.
- Enfermedades como Diabetes o Glaucoma.
- Degeneración macular.

Mientras que, la pérdida de la visión parcial puede variar según la causa:

- Presencia de cataratas, la visión puede estar nublada o borrosa y la luz brillante puede causar un molesto resplandor.
- Padecimiento de diabetes, la visión se torna borrosa, y la persona diabética suele tener dificultad para ver en la noche (en la gran mayoría de los casos).
- Personas con glaucoma, es posible que se produzca un estrechamiento concéntrico del campo visual y áreas de visión faltantes.
- Con la degeneración macular, podría aseverarse que la visión lateral es normal, no obstante, la visión central disminuye progresivamente.

Entre las causas no muy comunes que pueden presentarse están:

- Impedimento del flujo de los vasos sanguíneos.
- Dificultades derivadas del nacimiento prematuro (fibroplasia retro lenticular).
- Inconvenientes presentados en cirugías oftalmológicas.
- El llamado ojo vago o perezoso.
- Por Neuritis óptica

- Trastornos producidos por accidentes cerebrovasculares.
- A quienes padecen de retinitis.
- Los producidos por tumores.

2.2 Marco Conceptual Tecnologías de Información

2.2.1 Visión Artificial

"La visión artificial se ubica dentro de la inteligencia artificial, misma que basa su funcionamiento en emular la visión del ser humano con el fin de poder captar, procesar y analizar imágenes obtenidas a partir de dispositivos digitales". (Becker, 2015).

Según la Asociación de Imagenología Automatizada, cuyas siglas en inglés son AIA, (Automated Imaging Association), "la visión artificial abarca toda clase de aplicaciones, ya que realizan la función de la combinar sistemas de hardware y software para que los equipos realicen acciones dentro del procesamiento de imágenes".

Tanto en la academia como en la industria, en los gobiernes o en sus estrategias militares la visión artificial hace uso de una sola clase de algoritmos, independientemente a que las restricciones sean las mismas.

Componentes de la Visión Artificial

Entre los principales aspectos a considerar en un proceso de visión artificial es la iluminación, que permitirá obtener la mayor cantidad de información del objeto a ser analizado posteriormente. A continuación, se detallan los componentes que hacen parte del proceso:

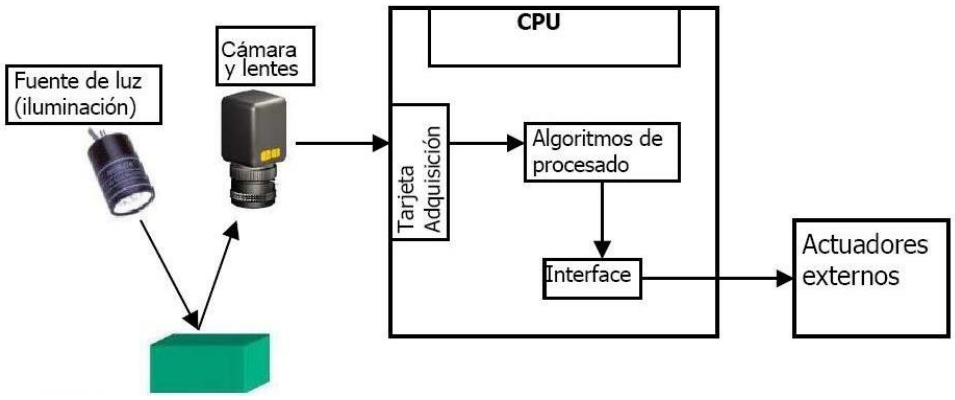


Figura 1. Componentes del prototipo de visión artificial

- **Iluminación**

La iluminación es uno de los aspectos más críticos de las aplicaciones de visión artificial, puesto que, si el objetivo no se ilumina adecuadamente, puede provocar pérdida de información y no será posible extraer las particularidades más transcendentales de la imagen a analizar. Lo cual, permitirá que la cámara obtenga una buena calidad sea procesada correctamente. La imagen que será objetivo de análisis debe contar con una buena iluminación.

- **Lentes**

La función de los lentes es capturar la imagen y determinar su resolución, la imagen se traslada de inmediato al sensor de imagen. La calidad del lente varía según las cualidades de sus mecanismos ópticos, mismos que definen la eficacia y el nivel de nitidez de una imagen. Gran parte de las cámaras con sistemas de visión artificial ofrecen dos clases de lentes: con repuestos y fijos. Para el desarrollo de este prototipo se utilizará una cámara USB de lente fija la misma que es parte de un sistema independiente y posee enfoque automático.

- **Sensor de imagen**

La capacidad de la cámara para capturar imágenes correctamente no depende únicamente del lente sino del sensor de imagen que se encuentra en el chip de la cámara. Esencialmente, la labor de un sensor es captar la luz para convertirla en una imagen digital. Vale acotar que la imagen es un cúmulo de píxeles. Si la luz es muy brillante producirá píxeles brillantes, caso contrario: si existe una presencia ligera de luz se producirán pixeles más sombríos.

- **Procesamiento de Visión**

Es un componente que sirve para obtener detalles de imágenes digitales, cuya labor se puede desempeñar mediante software de manera externa e interna. Los pasos a seguir son:

1) Se obtiene una imagen del sensor.
2) El software realiza una búsqueda de características específicas, a continuación; las medirá y las comparará con las descripciones detalladas.
3) Se debe comunicar los resultados.

- **Comunicación**

Se realiza a través de perifericos de entrada y salida, los datos son enviados a un dispositivo que registra la información. Los puntos de entrada y salida se conectan a un controlador lógico programable, y este utilizará estos datos para controlar los indicadores, en este caso la comunicación será con un dispositivo de audio, conectado directamente al computador.

2.2.2 Inteligencia

A lo largo de la historia, científicos, psicólogos, físicos, etc. Han propuesto varias definiciones sobre la inteligencia, originando verdaderos debates mundiales, entre éstos conceptos tenemos:

Freeman, manifiesta que las definiciones corrientes de la inteligencia se dividen en tres categorías:

- Las que dan mayor importancia al ajuste o adaptación del individuo a su medio total.
- Los aspectos especiales individuales para aprender o beneficiarse por medio de la experiencia.
- Las que subrayan la capacidad para llevar a cabo el pensamiento abstracto.

Para Howard Gardner, la inteligencia es un potencial bio-psicológico que procesa la información que activa uno o más marcos culturales para resolver problemas o crear productos de valor.

El psicólogo estadounidense Lewis Terman, describió a "La inteligencia como la capacidad para pensar abstractamente". Mientras que, Weschler, define a la inteligencia como: "La suma o capacidad global del individuo para actuar de un modo provisto de finalidad para pensar racionalmente y para tratar de enfrentarse de un modo eficaz con su medio".

Los autores Mira y López, aducen que: "Con la palabra inteligencia se designa actualmente la calidad o grado de eficiencia alcanzado por el trabajo mental destinado a la solución de dificultades adaptativas o al descubrimiento de nuevas relaciones de sentido entre los contenidos psíquicos"

Tales aseveraciones nos conducen a proponer que la inteligencia es la capacidad de resolver problemas para adaptarse ante cualquier situación.

2.2.3 Inteligencia Artificial

El término “Inteligencia Artificial” (IA), fue utilizado por primera vez en 1956 por el informático estadounidense John McCarthy. Hoy por hoy, es un término general que reúne todo, desde la automatización de procesos robóticos hasta la robótica actual. Ha ganado cada vez más importancia recientemente debido a las grandes cantidades de datos, o al aumento de velocidad, tamaño y variedad de datos que las empresas están recopilando.

La IA puede realizar tareas como identificar patrones en los datos de manera más

eficientemente que los seres humanos, lo cual permite a las grandes, medianas y pequeñas empresas, obtener más información sobre sus datos.

No obstante, la antesala para esta clase de inteligencia data del año de 1943 los científicos Walter Pitts y Warren McCulloch propusieron su propio modelo de "neurona artificial", en conjunto, presentaron las primeras hipótesis de lo que en la actualidad se conoce como "redes neuronales". Para los investigadores; "red neuronal" se traduce a un procedimiento inteligente que simula al sistema nervioso y a la estructura cerebral humana, sin embargo, la diferencia radica en su distribución. Un cerebro más complejo que cualquier red de neuronas artificiales, además las neuronas de origen artificial son unidades sencillas, en comparación con el sistema neuronal humano.

Sin embargo, tanto las neuronas de origen biológico, como las artificiales tienen la capacidad de entrelazar redes. La diferencia radica en que un sistema de neuronas artificiales procesa información de conexiones de entrada con otras neuronales artificiales, es decir que, activa las funciones de transferencia para permitir que la neurona en sus conexiones tenga una salida.

Margaret Rouse, manifiesta que la inteligencia artificial es la simulación de procesos de inteligencia humana por medio de máquinas, los más comunes se presentan en sistemas informáticos. Dichos procesos abarcan:

- Aprendizaje. - Es la adquisición de información y reglas para el uso de la información procesada.
- Razonamiento. - Utiliza las reglas para llegar a conclusiones aproximadas o definitivas y por último la autocorrección.

Las aplicaciones particulares de la AI incluyen sistemas expertos, reconocimiento de voz y visión artificial.

Primeros Modelos de Inteligencia Artificial

A través del tiempo, dentro del desarrollo de la inteligencia artificial varios modelos fueron presentados:

- Uno de los primeros estudios documentados en el área de esta clase de inteligencia se realizó a inicios de la década de los 40, Warren McCulloch y Walter Pitts dieron a conocer un modelo de neuronas artificiales.
- Dos décadas más tarde el investigador Lotfi Zadeh de la Universidad de Berkeley,

California ideó la teoría del conjunto difuso (" fuzzy sets", término que significa "difuso" como traducción del concepto "fuzzy") el cual fue difundido como “lógica difusa” para modelar el razonamiento humano muy a parte de la lógica clásica.

- En los años sesenta encontramos numerosos trabajos donde se hace uso de la *Teoría de Evolución de las especies*, cuyo autor fue Charles Darwin para diseñar algoritmos de optimización imitando la selección natural, siendo los primeros cimientos de lo que hoy se conoce como “Computación Evolutiva”, basada en el estudio del desarrollo de los algoritmos de evolución.
- Para la década de los 90´ se expande el término “inteligencia computacional” por James Bezdek y "soft computing" y por Lotfi Zadeh, creador de la lógica difusa.
- Los estudios antes mencionados tuvieron su origen bajo el nombre de “inteligencia computacional” o su terminología en inglés: "soft computing"

Inteligencia computacional (IC)

Está contemplada dentro de la inteligencia artificial. Su función diseñar sistemas informáticos inteligentes que intentan imitar la naturaleza y al razonamiento lingüístico humano.

Es la emulación del razonamiento lingüístico humano por medio de la teoría de los conjuntos difusos, donde las redes neuronales y los algoritmos de optimización bio-inspirados imitan al sistema nervioso.

Su importancia radica en el perfeccionamiento de sistemas inteligentes exitosos, incluidos los juegos y los sistemas de desarrollo cognitivo. En los últimos tiempos ha surgido una explosión de investigación sobre el aprendizaje profundo, especialmente en las redes neuronales convolucionales profundas.

Para el Instituto de Ingenieros Eléctricos y Electrónicos (IEEE), el término “inteligencia computacional” es explicado como: “una rama de la inteligencia artificial centrada en la teoría, el diseño y la aplicación de los modelos computacionales biológica y lingüísticamente inspirados, enfatizando las redes neuronales, los algoritmos evolutivos y bioinspirados, la lógica difusa, y los sistemas inteligentes híbridos en los cuales estos paradigmas están contenidos."

2.2.4 Redes Neuronales

La inteligencia artificial emplea redes neuronales con el propósito de emular las actividades que realiza el cerebro humano siendo una de las principales: el aprendizaje, que se efectúa a través de recopilación de información. La visión por computador utiliza este concepto para hacer que la máquina, realice procesos de reconocimiento, seguimiento y categorización de objetos mediante redes neuronales artificiales.

Redes Neuronales Artificiales (ANN)

Tienen su origen en la interacción de diversas partes simples que trabajan simultáneamente.

Se componen de redes artificiales que conectan los diversos procesadores ya que cuentan con algoritmos que se pueden ajustar sus neuronas (pesos) y de esta manera solventar algún determinado problema.

Cumplen las siguientes funciones: clasificar identificar, diagnosticar, optimizar y pronosticar.

El entrenamiento de una ANN, consiste en fomentar su adaptación dinámica a las necesidades de la aplicación readaptando y actualizando de manera constante los enlaces entre neuronas o llamados “pesos”.

En consecuencia, las redes neuronales artificiales pasan a ser algoritmos de gran potencia con los cuales se puede crear comportamientos inteligentes.

Una red neuronal artificial está compuesta por un gran número de conexiones definidas matemáticamente en la ecuación (1), emulando de esta forma una neurona biológica, su estructura se muestra en la siguiente formula:

$$\boldsymbol{z = f(b + \sum_{i=1}^{n}(wi + xi))} \quad 1$$

Donde:

$\boldsymbol{xi}$: entradas (valores entre 0 a 1)

$\boldsymbol{wi}$: pesos (valores entre 0 a 1)

$\boldsymbol{b}$: bias o sesgo

$\boldsymbol{f}$: Función de activación

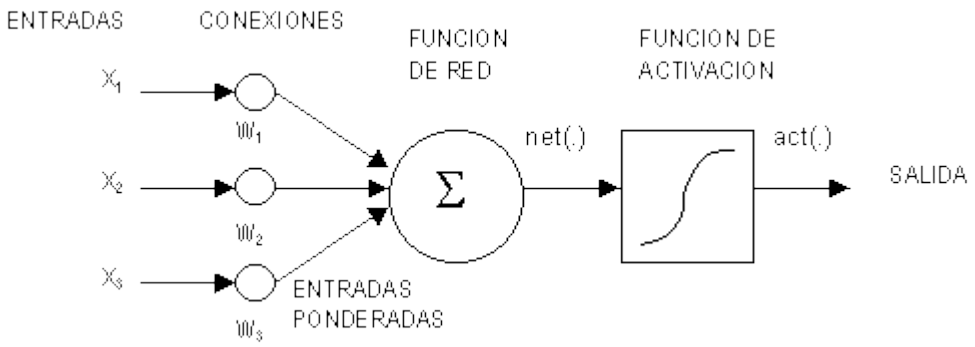


Figura 2. Estructura de una Neurona Artificial

Los parámetros mencionados cumplen una función específica; el peso, aumenta o disminuye el valor de entrada, mismo que toma valores de acuerdo con el nivel de importancia enfocado en el aprendizaje del modelo a implementar, inicialmente estos valores son aleatorios posteriormente durante el entrenamiento "training" poco a poco se van ajustando hasta lograr la tarea para la cual ha sido entrenado; el sesgo, permite ajustar el modelo a la función "activar" que básicamente hará que la señal de salida sea o no extendida al resto de las neuronas. Una red neuronal artificial está conforma por "n" número de neuronas conectadas entre sí, dispuestas en distintas capas tal como se puede observar en la Figura 2.

- Capa de Entrada: Recibe directamente los datos de entrada, en primera instancia los datos de entrenamiento de la red y posteriormente datos de prueba.
- Capa Oculta: Es responsable de manejar la labor de la red, tratando de mejorar cada vez su rendimiento y exactitud mejorando el aprendizaje. Una red puede contar con n capas ocultas.
- Capa Salida: Tiene la función de enviar resultados para los cuales la red ha sido entrenada.

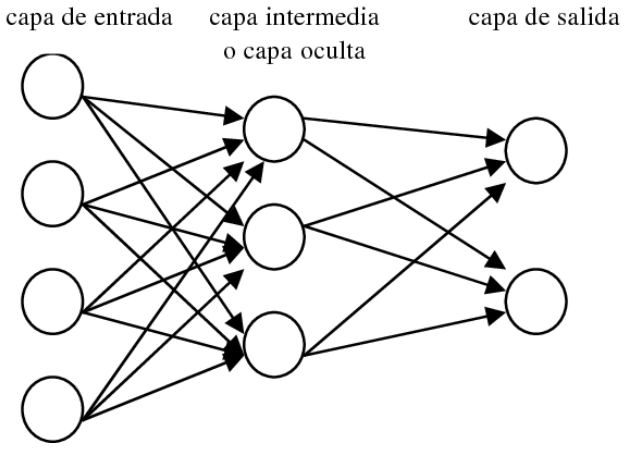


Figura 3. Estructura de una Red Neuronal Artificial

Una red neuronal convencional puede ser utilizada para implementar modelos de aprendizaje en cuanto a procesamiento de datos, modelización, predicciones numéricas etc., no obstante, para trabajar con imágenes que prácticamente son tratadas como matrices se requiere de redes neuronales especiales como son las redes neuronales convolucionales.

Red Neuronal Feed Forward

Conocidas también como layers, forman un vínculo de neuronas instauradas capa tras capa. Las cuales son:

- Capas de entrada
- Capas de salida
- Capas ocultas

Adquieren una función lineal, el eje es que la actividad de las neuronas, se transforme paralelamente con la aplicación, por tal motivo el resultado alcanzado será una función "no lineal", acoplada al nuevo patrón. Su forma de instalación es activar neurona tras neurona, esta acción hará posible la "no linealidad" a los resultados. Existen algunas funciones para activar la aplicación donde se manejan fundamentalmente las funciones como la Sigmoide o la RELU.

Redes Neuronales Convolucionales (CNN)

Corresponden a sistemas formados por redes neuronales de índole artificial, instauradas únicamente para el acoplamiento de la visión artificial.

Entre sus principales funciones están las de detectar y clasificar, así como de asimilar el grado que conlleva el perfil de una imagen.

Dicho sistema neuronal se halla dentro de campos receptores que se asemejan a la primera corteza visual del cerebro humano.

Esta clase de red neuronal tiene su origen a partir de aplicaciones que cuenten con varias capas, mediante una serie de filtros que emplean operadores convolucionales, durante la fase de entrenamiento se encargará de transformar y filtrar las imágenes de entrada. Tiene la misión de obtener los rasgos de dominio, para su cometido, trata de alcanzar un valor de pixeles más elevado.

Capas de redes neuronales convolucionales:

- De Entrada: Se ingresan las imágenes (sin procesar) para obtener información del entorno.

- De Convolución: Su fin es convertir la imagen mediante la aplicación de filtros para que ciertas particularidades sobresalgan del resto.
- De Activación: Cumplen con la función de activar la salida de las capas convolucionales, con el afán de adquirir una óptima configuración en el modelo a presentar, generan una "no linealidad".
- Capa de Pooling: Es filtro más delicado, encaminado a reducir el porte de la imagen digitalizada de entrada, gracias a este se obtienen rasgos independientes a la perspectiva o traslación de la imagen, además de reducir el tamaño de la misma, agilitando el procesamiento de imágenes en las siguientes capas. En esta fase el mapa de pixeles es dividido en sub-matrices y es donde se aplican las siguientes funciones:
- Maxpooling: Brindar el valor máximo del elemento analizado.
- Averagepooling: Mediante la sección de análisis se obtiene la media de todos los valores.

2.2.5 Proceso de Detección de Objetos

Algoritmos

La Real Academia Española (RAE), señala que la palabra algoritmo significa: "Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema", No obstante el origen de este término proviene de un matemático del siglo IX, llamado Al-Khwarizmi (sobrenombre de Mohamed ben Musa, quién estableció en los años 780-850 d.C, una solución metódica de problemas de álgebra y cálculo numérico mediante una lista bien definida, ordenada y finita de operaciones.

El primer algoritmo, conocido como tal, fue presentado por Charles Babbage en 1842, mediante su máquina analítica. Del mismo modo existe una versión de que el primer algoritmo informático a ser procesado por una máquina fue creado por la científica Ada Lovelace.

Hacia 1960 sesenta se dieron a conocer los primeros modelos de algoritmos de optimización, cuyo objetivo era resolver problemas imitando a la naturaleza. Hasta llegar a nuestros a nuestra era contemporánea donde los algoritmos ocupan un gran citial en todo lo que respecta a la Informática.

Por otro lado, para el catedrático español Ricardo Peña, algoritmo es "Conjunto de reglas que, aplicada sistemáticamente a unos datos de entrada apropiados, resuelven un problema en un numero finito de pasos elementales*".* Sin embargo, vale la pena acotar que los algoritmos no solo deben estar asociados a la matemática, no es de uso exclusivo de ninguna ciencia.

No obstante, los algoritmos de detección de objetos no ejecutan una misión tan sencilla

como parece, para entender y asumir su complejidad, este estudio se basará en que los seres humanos tienen la capacidad de ver una fotografía y de inmediato reconocer cualquier objeto por simple inspección; pueden discernir el tamaño forma y color de los objetos.

En cambio, para detectar imágenes con algoritmos de "Machine Learning" una (CNN) debe diferenciar una cantidad específica de objetos. A demás los Algoritmos de Detección de Objetos, comprenden una tecnología de ordenador similar a la visión artificial y el procesamiento de imagen que pretende detectar casos de objetos semánticos de una cierta clase (como personas, casas, edificios o automóviles, etc.) en vídeos e imágenes digitales.

A continuación, se realiza el detalle del desarrollo del algoritmo de reconocimiento de objetos, mismo que está jerarquizado por etapas, en las que se manejan técnicas para digitalizar imágenes y para ejecutar procesos de Machine y Deep Learning sean para entretener o implementar modelos de redes neuronales designadas.

Técnicas:

Digitalización de Imágenes

Las imágenes están compuestas por pixeles (puntos que componen una imagen) los pixeles también son conocidos como "unidades mínimas de medición de una imagen en digital". Cada valor que posee un pixel hace referencia a una cromática primaria digital conocida como RGB siglas que traducidas del inglés al español quieren decir: Red = rojo, Green = verde y Blue = azul.

El proceso cumple el trabajo de digitalizar imágenes también comprende la labor de buscar secuencias de pixeles que se parezcan al objeto que se pretende identificar.

A través del uso de técnicas como Machine Learning y Deep Learning algoritmo de identificación, puede cumplir con el trabajo de clasificar y localizar objetos. Es preciso añadir que la identificación de objetos es muy compleja, por lo tanto se ubica dentro del aprendizaje automático.

Aprendizaje automático

Está enfocado en el perfeccionamiento de programas que permitan el rápido acceso y generalización de datos, además es considerado como una derivación de la inteligencia artificial, en el cual sistema no está previamente programado, se basa en su empirismo, donde los sistemas tienen la capacidad de desarrollar aprendizajes de manera automática.

Por consiguiente, el aprendizaje automático corresponde al estudio de instrumentos

informáticos que llegan a tomar decisiones por sí solos. Por ejemplo: Machine Learning combina varias ciencias como: matemática, programación, lógica, estadística, etc.

Existen varias clases de aprendizaje, para la elaboración del presente trabajo se utilizará el:

Aprendizaje automático supervisado

Los algoritmos de aprendizaje realizan las debidas comparaciones entre las predicciones de correcta para localizar errores y repararlos a tiempo. Este tipo de aprendizaje se obtiene tras el proceso de datos de entrenamiento.

Posee la funcionalidad de utilizar experiencias adquiridas durante el período de entrenamiento para poder generar nuevas deducciones. Las mismas que dependerán en mayor proporción del compendio de datos obtenidos luego del entreno de los algoritmos.

Actualmente, no existe un modelo de aprendizaje 100% confiable, razón por la cual, es preciso estudiar las aplicaciones para implementar el modelo requerido.

El algoritmo elegido para sr usado en la fase de detección está formado por por una CNN y una Red Layers las cuales permitirán al sistema la detección, localización y clasificación de objetos

2.2.6 Hardware

Computador de placa única (SBC)

Es un tipo de computadora que cuenta con todos sus componentes integrados en una sola placa; como memoria RAM, microprocesador, ranura para la tarjeta de memoria que ejecutará las funciones del disco duro, al mismo tiempo tienen entradas y salidas para la conexión de periféricos.

Conforme a las características del proyecto se realizará el procesamiento en un equipo Raspberry Pi 4 Model B, ya que principalmente en su relación rendimiento-precio es el más indicado para la ejecución del software.

Raspberry Pi 4 Modelo B

Raspberry Pi surge en 2006 como una iniciativa de bajo precio para promover la experimentación con la programación desde edades tempranas, su misión es fomentar la enseñanza de las ciencias de la Informática a los niños.

Siendo su sistema operativo oficial una versión adaptada de Debian, denominada

Raspberry Pi OS, no obstante, permite usar otros sistemas operativos, incluido una versión de Windows 10. En todas sus versiones, incluye un procesador Broadcom, memoria RAM, GPU, puertos USB, HDMI, Ethernet, 40 pines GPIO y un conector para cámara.

Raspberry Pi proporciona un entorno de aprendizaje a un bajo costo, ideal para implementar proyectos enmarcados a la visión artificial. En otra instancia, se deben tener claras sus posibles desventajas, como por ejemplo la potencia reducida de su GPU, memoria RAM y del microprocesador. Sin embargo el mayor problema presenta en su ciclo de entrenamiento. Por tal razón, se ve la necesidad de cumplir con el transcurso de entrenamiento del modelo con un ordenador con más potencia.

Especificaciones

Procesador	Broadcom BCM2711, quad-core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz
Memoria	2GB or 4GB LPDDR4
Conectividad	2.4 GHz and 5.0 GHz IEEE 802.11b/g/n/ac wireless LAN, Bluetooth 5.0, BLE Gigabit Ethernet 2 × USB 3.0 puertos 2 × USB 2.0 puertos.
GPIO	Standard 40-pin GPIO header
Video y Sonido	9 micro HDMI puertos 2-lane MIPI DSI 2-lane MIPI CSI camara 4-pole audio y video puertos
SD Card	Micro SD tarjeta. SO y memoria
Corriente carga	5V DC via USB-C 5V DC via GPIO Power over Ethernet (PoE)–enabled (requires separate PoE HAT)

Tabla 1. Especificaciones Raspberry PI 4 B

Cámaras USB

Su función es capturar la imagen proyectada en el sensor para poder transmitirla electrónicamente y ser procesada e interpretada. Cuyo caso, el procesamiento de imágenes se realizará en la Raspberry.

Conforme al procesamiento para visión artificial, es obligatorio contar con ciertas características específicas, como la rapidez del obturador, además del parámetro pixeles ya que es necesario contar con la mayor cantidad de detalles al momento de su procesamiento.

En síntesis, una cámara USB, incorpora los elementos y componentes tradicionales de las cámaras clásicas, como los sensores y la electrónica asociada a la captura de la imagen; conjuntamente engloban un procesador, memoria y sistemas de comunicación con el exterior a través de buses.

Entre las principales ventajas de las cámaras USB están:

- Portabilidad: Las cámaras USB son menos propensos al sobrecalentamiento y contienen cubiertas más grandes que admiten que el calor se disperse más rápidamente. Los dispositivos son portátiles y se pueden desconectar de un dispositivo y conectarlo a otro sin requerir procesos de instalación extensos.
- Compatibilidad: Las cámaras USB son compatibles con una variedad de plataformas, incluido Windows, Mac, Linux e incluso sistemas operativos de para consolas de video juegos.

Dispositivos de audio

Comúnmente conocidos como controladores de audio, son aquellos aparatos tecnológicos que tienen la labor de procesar directa o indirectamente señales de ondas auditivas, también pueden generarlas o transformarlas.

Tienen la función de las frecuencias eléctricas convertir en energía mecánica, y esta a su vez, puede ser convertida en energía acústica. El sonido se propaga a través del aire donde circulan las ondas sonoras. A su vez llegan al órgano auditivo de la persona a manera de propulsiones nerviosas que estimulan la masa encefálica.

Entre los principales dispositivos de audio están:

- Sistema de altavoz. - Es un conductor electro-acústico creado para reproducir sonidos. Está compuesto de altavoces que forman una pantalla acústica.
- Tarjeta de sonido. - Se trata de dispositivo hardware de reducido tamaño que permite al ordenador procesar sonidos de entrada y salida.
- Dolby. - La cualidad que diferencia a este dispositivo de otros es que actúa únicamente sobre los sonidos de bajo volumen, es decir los que se ven más afectados por el ruido.
- Surround.- Se trata del uso de múltiples canales de audio que sirven para provocar efectos envolventes a los usuarios. Esta tecnología se ha instalado en los hogares como parte fundamental de los sistemas de cine en casa.
- PIP. – Su función es la de manejar las partes de la pantalla por separado para poder

observar canales distintos y grabarlos.

- THX. - Realiza una inspección de calidad, no sirve para grabar y no es un formato de sonido, puesto que, todos los sistemas tanto audio digitales como analógicos pueden ser escuchados en THX.
- S/PDIF. – sirve para transmitir señales de videos digitales.

2.2.7 Software

Considerando la importancia de una elección correcta en cuanto al software se consideraron aspectos clave como el rendimiento y las características del hardware que se va a utilizar, es por ello que la ejecución del código se realizará utilizando el sistema operativo Raspbian en su versión Buster, el lenguaje de programación a utilizarse es el de Python.

Python.

Python es un lenguaje de programación de uso libre, además de ser interpretativo y de multipropósito fácil de leer y comprender, y pone más énfasis en la legibilidad del código para facilitar la comprensión de la sintaxis. El código sencillo e implementado de Phyton hace que sea fácil de aprender, tanto para principiantes como para aquellos que dominan el lenguaje de programación. Varias plataformas de sistemas operativos y requisitos de desarrollo de software pueden usar Python.

El lenguaje de programación Python se define como un lenguaje que combina capacidades, que con una sintaxis de código muy clara y también viene con una gran funcionalidad de biblioteca estándar integral. Python es una herramienta gratuita, no hay restricciones sobre cómo distribuirlo. Equipado con gestión de memoria automática, Python es un lenguaje de programación dinámico, cuya programación es fácil de entender debido a la forma clara del lenguaje utilizado (Aviolita, 2019).

Para la ejecución del presente trabajo de investigación se hará uso de las siguientes librerías:

Numpy

Pertenece a Python, se especializa en cálculo numérico y el análisis de datos, especialmente proyectado para un gran volumen de datos. La ventaja de Numpy frente a las listas predefinidas en Python es que el procesamiento de los arrays, actúa con más rapidez que las listas, lo cual lo hace ideal para el procesamiento de matrices con grandes dimensiones.

OS

Este módulo provee una manera versátil de usar funcionalidades, como la administración y creación de archivos que interactúan de forma directa con los sistemas operativos.

Proporciona detalles acerca del entorno y permiten la manipulación de la estructura de listas de lectura y redacción.

Tensorflow Lite

Es un marco de aprendizaje profundo multiplataforma y de código abierto que convierte un modelo previamente entrenado en TensorFlow en un formato especial que se puede optimizar para la velocidad o el almacenamiento. Esta clase de formato fácilmente se puede implementar en dispositivos móviles o dispositivos integrados basados en Linux como Raspberry Pi.

Tensorflow Lite Converter convierte un modelo de Tensorflow en un archivo de búfer plano de Tensorflow Lite (.tflite)

Tensorflow Lite Model Maker: La biblioteca TensorFlow Lite Model Maker agilita el proceso de entrenamiento de un modelo de TensorFlow Lite con un conjunto de datos personalizado. Utiliza el aprendizaje por transferencia para reducir la cantidad de datos de entrenamiento requeridos y acortar el tiempo de entreno.

OpenCV

Sus siglas significan: Open Computer Vision que en español se traducen a Visión Artificial Abierta. La primera versión fue presentada en 1999, a través de los años ha sido utilizada en una gran cantidad de aplicaciones, hasta la actualidad se la sigue mencionando como la biblioteca de visión artificial más popular, al ser una biblioteca de programación informática de código abierto desarrollada para admitir aplicaciones que utilizan visión artificial.

OpenCV abarca cientos de funciones que apoyan la captura, análisis y manipulación de información visual alimentada a una computadora por cámaras web, archivos de video u otros tipos de dispositivos.

Algunos ejemplos de aplicaciones de OpenCV son: Detección de movimiento, reconocimiento de objetos, reconstrucción 3D.

LabelImg

Es una herramienta de etiquetado y anotación de imágenes liviana y fácil de usar para etiquetar cuadros delimitadores de objetos en imágenes.

Posee una interfaz gráfica completa y se encuentra escrita en Python y Qt. Su mejor característica es que la información de anotación se puede convertir directamente en un archivo XML, que es el mismo que el XML utilizado por PASCAL VOC e ImageNet.

Google Text to Speech (gTTS)

Peretenece a Python y una herramienta CLI para interactuar con el API de texto a voz de Google Translate. gTTS es una herramienta que convierte el texto ingresado en audio que se puede guardar como un archivo mp3. La variable de texto es una cadena utilizada para almacenar la entrada del beneficiario, en cuyo caso, la entrada será proporcionada por el módulo de detección de objetos y posteriormente se convertirá en una salida de audio.

Puede ser manejada por aplicaciones como Google Play Books para leer libros en voz alta, por Google Translate para leer traducciones en voz alta, brindando información ventajosa sobre la pronunciación de las palabras, por Google Talkback y otras apps basadas en la accesibilidad de comentarios hablados, así como como por otras aplicaciones. Para su uso los usuarios deben instalar datos de voz para cada idioma.

Mpg321

Mpg321 es un proyecto inventado y escrito por Joe Drew. Es un reproductor de mp3 por línea de comandos que utiliza la biblioteca de decodificación local. Fue desarrollado para reemplazar el reproductor mpg123 que no es libre.

A más de ser muy popular entre los usuarios, es utilizada para frontends, como reproductor de mp3 y como decodificador de archivos mp3 a wave (principalmente para usar con software de grabación de CD).

2.2.8 Metodología

"La metodología constituye una de las fases más importantes de un determinado proyecto, lo cual conlleva a la seleccionar técnicas y métodos concretos para alcanzar y cumplir con los objetivos establecidos". (Maida & Pacienzia, 2015).

En otra instancia: para obtener un software de calidad se debe utilizar una metodología que permita especificar los requerimientos del usuario, tales como: diseño, experimentación e implementación del producto.

Dentro de estas metodologías de desarrollo se destacan los métodos ágiles, mismos que, a lo largo del transcurso de los años han servido de gran aporte para el perfeccionamiento de aplicaciones.

Para la ejecución de este proyecto en particular y al tratarse de un prototipo se emplean técnicas de consulta y experimentación que permitan el desarrollo e implementación del prototipo por avances, para ello se establecen los siguientes periodos:

- El primer periodo consiste en la recopilación de requerimientos del prototipo, en el cual se detallan las características funcionales del mismo, además se establecen los objetos que serán reconocidos por el prototipo y las características del mismo.
- El segundo periodo consiste en el diseño del prototipo, y la descripción del entrenamiento del modelo de acuerdo a la identificación de características y objetivos.
- Finalmente, durante la fase de implementación del sistema, se procederá a la unificación del sistema en lo que tiene que ver con el hardware y software.

Otra de las herramientas utilizadas son las encuestas, las mismas que constituyen un procedimiento dentro de la investigación cuantitativa en la que el investigador recopila información mediante el cuestionario previamente diseñado, sin modificar el entorno ni el fenómeno donde se recoge la información ya sea para entregarlo en forma de tríptico, gráfica, tabla o escrita. A partir de los resultados conseguidos se puede establecer varios de los parámetros necesarios para el diseño del prototipo

Capítulo 3: Desarrollo del Prototipo

En el presente capítulo se detalla el desarrollo del prototipo, tomando como referencia principal cada uno de los objetivos específicos planteados para la realización de la presente tesis. La descripción parte desde su fase de identificación de objetivos, describiendo el diseño del mismo y finalmente la implementación del hardware y software.

3.1 Identificación de objetivos de detección con el uso de software libre de aprendizaje automático y bibliotecas de visión artificial

Acorde a la definición de prototipo este dispositivo no contará con todas las características y/o llevará a cabo la totalidad de funciones requeridas para un dispositivo final; incluirá elementos suficientes para permitir la evaluación e identificar aquellas características que agregarán valor al dispositivo y los posibles cambios que requiere el mismo. En este caso este dispositivo permitirá verificar la factibilidad del diseño propuesto.

El dispositivo, inicia al encender la Raspberry con lo cual, se enciende la cámara y de forma automática la cámara empieza a recolectar información en tiempo real en forma de imágenes en el entorno de la persona.

La detección de objetos se lleva a cabo una vez que el beneficiario presiona el botón ubicado en la Raspberry, en consecuencia, se activan las respectivas funciones para permitir el procesamiento y detección de imágenes para finalizar con la salida del audio hacia el oído del usuario.

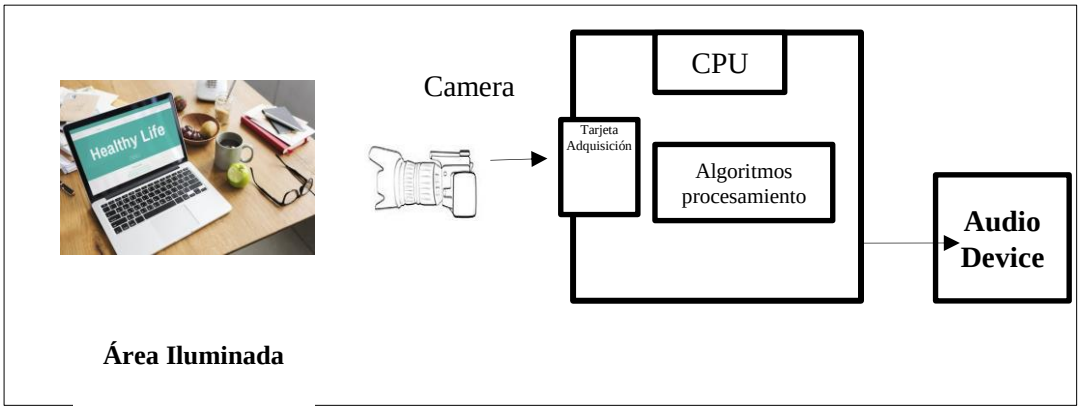


Figura 4. Diagrama general de funcionamiento

3.1.1 Características principales del prototipo

Para poder determinar las características del prototipo es preciso entender las necesidades del usuario en términos de los atributos del producto, para que los mismos se traduzcan en especificaciones apropiadas a las cualidades de ingeniería.

Dichas necesidades, se pueden analizar a través de los resultados de opiniones directas con los posibles involucrados en el uso del dispositivo.

Inicialmente se determinará el grupo objetivo, principalmente se incluyen a todas las personas con ceguera total, mientras que por otro lado su uso se destinara a personas como cierto nivel de discapacidad visual. Se pueden recopilar las siguientes características.

- Hombres y mujeres a partir de 10 años.
- Personas con nivel de ceguera baja y alta.

De estos grupos se elegirá a la mayor cantidad de personas posibles para determinar los objetos de detección.

Finalmente se examinarán cuidadosamente las características que requieren los usuarios en el prototipo y los requerimientos de este; con el fin de obtener los objetivos de detección, así como las características claves del prototipo.

3.1.2 Análisis e interpretación de encuestas

Con el propósito de poder obtener la mayor cantidad de información posible, acerca de las características del prototipo y los principales objetos que el mismo debe reconocer, se adquiere el valor de la muestra considerando los siguientes aspectos, el nivel de confianza de la información proporcionada se estima en 96%, mientras que las probabilidades a favor y en contra serán del 50%, al tratarse de un prototipo el error se calcula en un 20%, ya que se prevén futuras modificaciones de acuerdo al desarrollo del dispositivo.

$$n = \frac{z^2 * p * q}{e^2} = \frac{1.96^2 * 0.5 * 0.5}{0.2^2}$$

$$n = 24.01$$

El valor calculado de la muestra es de 24, por motivos prácticos la encuesta se realizará a 30 personas con diferentes niveles de discapacidad visual, incluidas aquellas con ceguera y enfermedades relacionadas a la vista, estas encuestas se aplican de forma directa obteniendo la información de forma verbal.

Edad

El rango de edad de la población encuestada oscila entre 10 a 45 años, edades que abarcan personas en etapa de niñez hasta usuario en periodo de adultez.

1.- Como le gustaría que fuese un dispositivo de visión artificial:		
Gafas de uso común		85%
Pulsera con cámara		2%
Dispositivo colgante en el pecho del usuario		3%
Aplicación disponible desde el celular		10%
TOTAL		100%

Tabla 2. Pregunta 1 Encuesta Identificación de objetivos y requerimientos del usuario

Posterior a la aplicación de las encuestas se obtuvo como resultado que el 85% de las personas no videntes les gustaría que las gafas sean de uso común. Mientras que el 2% desearía un modelo en base a pulseras con cámara. El 3% respondió que desearía un dispositivo que cuelgue en el pecho y el 10% está de acuerdo que el dispositivo fuese diseñado a través de una aplicación de celular. Por ende, podemos deducir que la mayoría de los encuestados requiere que el dispositivo de visión artificial sea mediante gafas de uso común.

2.- En cuanto al precio, ¿cuál es el rango que estaría dispuesto a pagar?		
$50 - $100		98%
$100 - $200		1%
$200 – más		1%
TOTAL		100%

Tabla 3. Pregunta 2 Encuesta Identificación de objetivos y requerimientos del usuario

Del total de encuestas realizadas el 98% de encuestados considera que, están dispuestos a pagar de $50 a $100, Mientras que el 1 % manifestó que pagaría de $100 a $200, el 1% restante opinó que pagaría $200 en adelante. Por lo tanto, el precio debe estar contemplado dentro del alcance del presupuesto de la mayoría, basándose en un rango de 50 a 100 dólares.

¿Cómo prefiere que sea su modo de uso?		
Botón pulsador		75%
Varios botones de acciones		15%
Acción por voz		10%
TOTAL		100%

Tabla 4. Pregunta 3 Encuesta Identificación de objetivos y requerimientos del usuario

Según los resultados de las encuestas el 75% de los ciudadanos no videntes respondió que prefiere que el modo de uso sea a través de un botón pulsador, en contraposición al 15% que se acomodaría a varios botones de acciones, y por último el 10% manifestó que prefiere el uso

mediante acción de voz. Resultados de los cuales se puede inferir que la gran mayoría de usuarios prefieren que el producto debe ofrecer un modo de uso que sea a través de un botón pulsado.

4.- ¿Cuáles son las características que considera más importantes de acuerdo a la batería?		
Dispositivo autónomo		0%
Baterías recargables		30%
Baterías de larga duración		40%
Costo accesible		18%
Tiempo de carga		12%
TOTAL		100%

Tabla 5. Pregunta 4 Encuesta Identificación de objetivos y requerimientos del usuario

La población encuestada, en cuanto a las características que considera más importantes conforme a la batería se manifestó de la siguiente manera; el 0% eligió la opción de dispositivo autónomo, un 30% eligió baterías recargables, otro 40% escogió la opción de baterías que cuenten con larga duración. Además, un 18% consideró importante que el costo sea accesible. Y el 12% corresponde a que tiene relevancia el tiempo de carga. Por consiguiente, para el diseño, implementación y difusión del dispositivo se tomarán en cuenta las siguientes características: la carga de energía será a través de baterías recargables y de larga duración, así como su tiempo de carga y el costo accesible. Se descarta el aspecto de que el dispositivo sea autónomo.

5.- ¿Cuál es la relación que considera más importante para el dispositivo?		
Alta velocidad de respuesta – Baja precisión de reconocimiento		62%
Baja velocidad de respuesta – Alta precisión de reconocimiento		38%
TOTAL		100%

Tabla 6. Pregunta 5 Encuesta Identificación de objetivos y requerimientos del usuario

El 68% de las personas encuestadas determinaron que la relación que consideran más importante para el dispositivo es la de Alta velocidad de respuesta – Baja precisión de reconocimiento. Por su parte el 32% opinó lo contrario. Es decir que, a través de estos resultados se pudo concluir que fue muy reñida la decisión de los usuarios, sin embargo, la relación Alta velocidad de respuesta – Baja precisión de reconocimiento fue la alternativa más opcionada por quienes respondieron esta pregunta.

6.- Seleccione las características considera más importantes para el dispositivo		
Autónomas		1%
Recargable		28%
Interfaz de usuario sencilla		2%
Fácil transportación		7%
Alta precisión		10%
Fácil mantenimiento		22%
Económicas		30%
TOTAL		100%

Tabla 7. Pregunta 6 Encuesta Identificación de objetivos y requerimientos del usuario

Entre las características más importantes, referentes al dispositivo, los encuestados respondieron de la siguiente manera: el 1% opinó que está el hecho de que el dispositivo sea autónomo, un 28% asume que es importante que sea recargable, el 2% optó porque tenga una interfaz sencilla, el 10% corresponde a que sea de alta precisión, Mientras que el 22% manifiesta que debe ser de fácil mantenimiento, y en última instancia, un 30% solicita que el producto sea económico. Por lo tanto, las características que serán tomadas en cuenta para el dispositivo son: Que sea recargable, de fácil mantenimiento y económicas, relegando a los demás aspectos como: autonomía, interfaz de usuario sencilla, fácil transportación y de alta precisión a un segundo plano.

7.- ¿Cuáles de las siguientes considera las categorías más importantes para el reconocimiento de objetos?		
Medicinas		3%
Animales		14%
Dinero		3%
Personas		22%
Dispositivos electrónicos		19%
Alimentos		18%
Utensilios		15%
Maquillajes		1%
Muebles		3%
Objetos de aseo personal		2%
TOTAL		100%

Tabla 8. Pregunta 7 Encuesta Identificación de objetivos y requerimientos del usuario

A través de los resultados se obtuvo: medicinas 3%, animales 14%, dinero 3%, personas 22%, dispositivos electrónicos 19%, alimentos 18%, utensilios15%, maquillajes 1% muebles 3% y objetos de aseo personal 2%. Se pudo concluir que entre las categorías para el reconocimiento

de objetos que los usuarios consideran más importantes están: los animales, las personas, dispositivos electrónicos, y alimentos, utensilios.

8.- Cuáles de los siguientes objetos considera necesarios para el reconocimiento de objetos		
TV		5%
Manija		0%
Celular		12%
Frutas y vegetales		14%
Laptop		5%
Monedas		2%
Cama		0%
Mochila		5%
Zapatos		2%
Bastón		0%
Silla		2%
Jabón de baño		4%
Botellas		1%
Billetes		3%
Puerta		5%
Cubiertos		10%
Perros		8%
Personas		15%
Ropa		5%
Esfero		2%
TOTAL		100%

Tabla 9. Pregunta 8 Encuesta Identificación de objetivos y requerimientos del usuario

Según los resultados de las encuestas, los usuarios seleccionaron como objetos considera necesarios a los siguientes: televisión 5%, manija 0%, celular 12% frutas y vegetales 14%, laptop 5%, monedas 2%, cama 0%, mochila 5%, zapatos 2% bastón 0%, silla 2%, jabón de baño 4% botellas 1%, billetes 3%, puerta 5%, cubiertos 10%, perro 8%, personas 15%, ropa 5%, y esfero 2% . En consecuencia, el grupo encuestado eligió los objetos más necesarios para el reconocimiento de objetos a: frutas y vegetales, cubiertos, perros y personas. Mismo que serán tomados en cuenta para el diseño del presente dispositivo.

3.1.3 Requerimientos del usuario

Para obtener los requerimientos del usuario, se realizó encuestas de manera verbal a personas no videntes, a quienes se les consultó cómo desearían que fuera el producto y las características que quisieran que tenga para luego transmitirles la idea de cómo diseñar un prototipo en su beneficio.

IDEAS DEL USUARIO		REQUISITOS DEL USUARIO
1	**Gafas de uso común, autónomas**	Las gafas deben contar con un sistema inteligente capaz de reconocer objetos en el entorno de una persona.
2	**Económico**	El prototipo debe contar con partes que resulten accesibles de acuerdo al costo y beneficio.
3	**Fácil de utilizar**	Fácil manejo y comodidad al manipular
4	**Recargable**	Baterías recargables y de larga duración
5	**Batería de larga duración**	
6	**Rápido**	El dispositivo requiere respuestas inmediatas.
7	**Silencioso**	Silencioso
8	**Accesorios intercambiables**	Capacidad de desmontaje sencilla
9	**Fácil mantenimiento**	
10	**Fácil transportación**	
11	**Resistente**	Estructura resistente, de tamaño manejable y ergonómica para el usuario
12	**Apariencia**	
13	**Seguridad para el usuario**	Seguro

Tabla 10. Requerimientos del Usuario

3.1.4 Características Técnicas Obtenidas

Una vez establecidos los requerimientos del usuario, se procede a enlistar las características técnicas con el fin de dar una solución a cada requisito establecido por el beneficiario.

REQUERIMIENTOS DEL USUARIO		CARACTERÍSTICA TÉCNICA
1	Las gafas deben contar con un sistema inteligente capaz de reconocer objetos en el entorno de una persona.	**Algoritmo de reconocimiento de objetos**
2	El prototipo debe contar con piezas que resulten accesibles de acuerdo al costo y beneficio.	**Costo de elementos**
3	Fácil manejo, comodidad al manipular	**Interfaz de usuario sencilla**
4	Baterías recargables y de larga duración	**Consumo de energía**
5	El dispositivo requerirá de respuestas inmediatas.	**Velocidad de procesamiento**
6	Silencioso	**Tipo de computador y ventiladores**
7	Capacidad de desmontaje sencilla	**Estructura**

8	Estructura resistente, de tamaño manejable y ergonómica para el usuario	**Diseño ergonómico**

Tabla 11. Características técnicas del prototipo

Identificación de objetivos

A partir de la información recabada de los principales involucrados en el uso del dispositivo es indispensable el reconocimiento de objetos, de uso diario, tales como utensilios de comida, entre ellos, cubiertos, vasos, platos, tazas. Por otra parte, el usuario está en contacto continuamente con otras personas e incluso animales; dispositivos electrónicos es otra categoría importante para las personas no videntes. Adicionalmente, comida, alimentos entre ellos, frutas, verduras, enlatados y botellas.

Para la elaboración del prototipo se debe considerar la construcción de un equipo autónomo, con un funcionamiento básico y específico para el usuario, de tal forma que se especifiquen los siguientes requisitos:

- Los datos serán procesados por el computador para luego ser enviados como información específica para el usuario.
- El prototipo tendrá la capacidad de reconocer objetos en el entorno de una persona.
- El dispositivo contara con suficiente autonomía para trabajar durante un periodo de tiempo con el uso de batería recargable.

En relación a los objetos, se definieron las siguientes categorías y con el fin de determinar la funcionalidad del prototipo se trabajará con los siguientes objetos.

Categoría	**Objeto**
Dispositivos electrónicos	Celular
Alimentos	Manzana
Utensilios	Botella
	Taza
Animales	Perro
	Gato
Personas	Hombre
	Mujer

Tabla 12. Identificación de objetos

3.2 Diseño de un prototipo de gafas mediante el uso de una cámara, un dispositivo de audio y el uso de un computador de placa reducida para beneficio de personas no videntes.

En la Figura 5 se muestra el diagrama esquemático del prototipo, donde se expone cada dispositivo externo usado en la Raspberry PI. El uso del dispositivo se lo realizara de acuerdo con los requerimientos obtenidos por los usuarios. Por su parte el sistema se encarga de procesar la información del entorno de la persona por medio de capturas de las imágenes, para que pasen al proceso de reconocimiento y finalmente, la transmisión de los avisos de audio cumple la función de identificar los objetos cercanos al usuario.

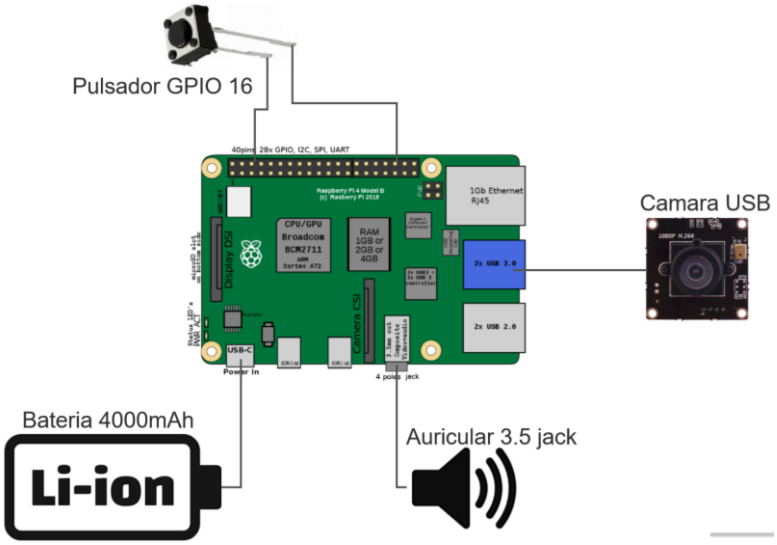


Figura 5. Esquemático conexión Raspberry PI a los dispositivos externos

Cuando se pulsa el botón que permite la función de reconocimiento, se reproduce automáticamente un mensaje de voz con el nombre del objeto que fue reconocido; una vez que se presiona el botón el sistema realiza una captura de imagen, la cual que es procesada por algoritmos de reconocimiento de objetos y finalmente se toman los parámetros del algoritmo para ser reproducidos por audio al oído del usuario. En la Figura 6 se puede observar el diagrama de flujo general del funcionamiento del prototipo.

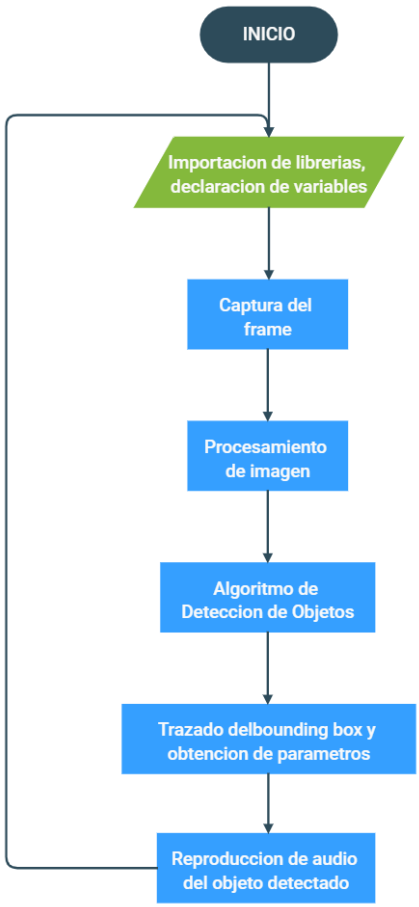


Figura 6. Diagrama general de flujo

3.2.1 Proceso de reconocimiento de objetos

La Figura 7 muestra el proceso especifico de reconocimiento de objetos, una vez que el obturador obtiene una imagen en tiempo real del entorno de la persona, inicia el proceso "Reconocimiento", la red neuronal inicia su construcción en base a los pesos obtenidos, a continuación, se realiza la filtración, procesamiento y reconocimiento de objetos en la imagen.

Con los parámetros obtenidos de la imagen, y con la cantidad de objetos detectados se procede a la recuperación de audio de la base de datos establecida para la posterior reproducción de audios.

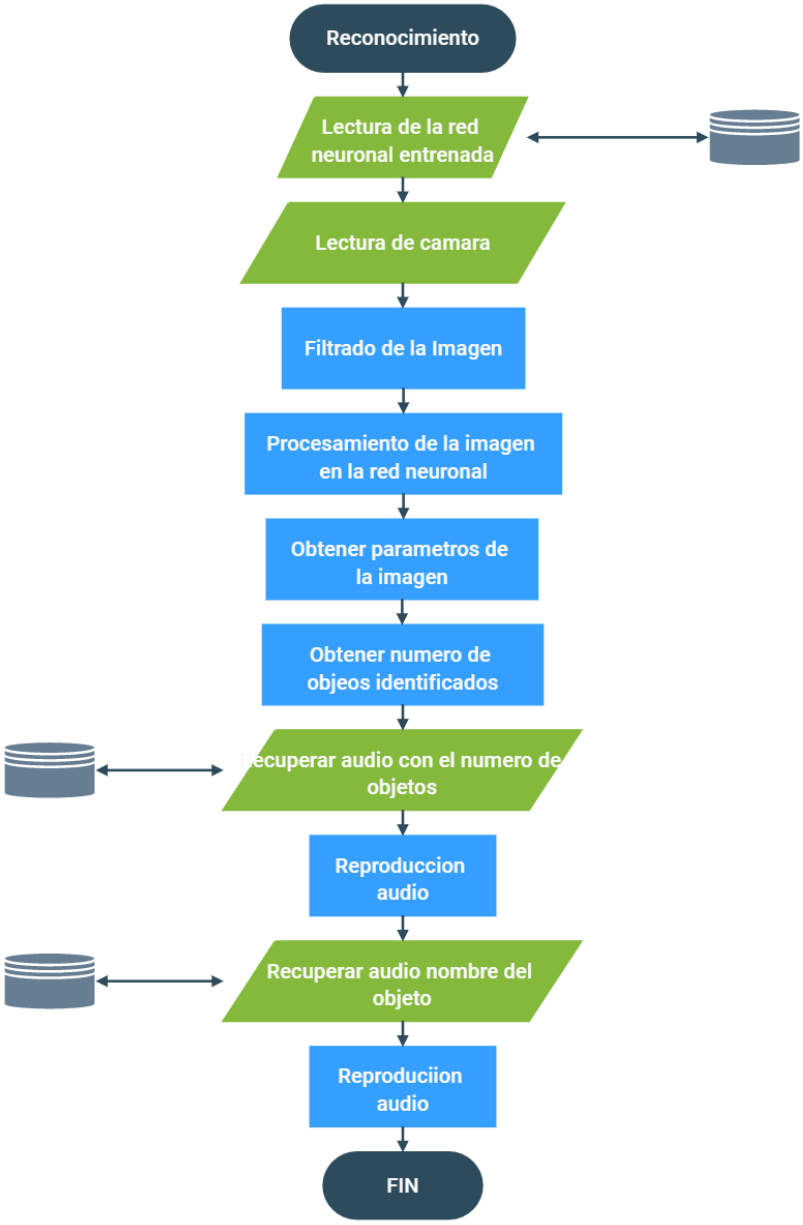


Figura 7. Diagrama de flujo del reconocimiento de objetos.

3.2.2 Entrenamiento del modelo de detección (personalizado)

Por tratarse de un prototipo se deben realizar pruebas con distintas clases de objetos, como parte del proceso es preciso dividir los archivos en conjuntos de imágenes de entrenamiento y confirmación por tal razón, se trabaja con alrededor de 200 imágenes de cada objeto; a continuación, se presenta el listado de los objetos incluidos en el modelo entrenado.

Celular	Taza	Mujer
Manzana	Perro	Hombre
Botella	Gato	

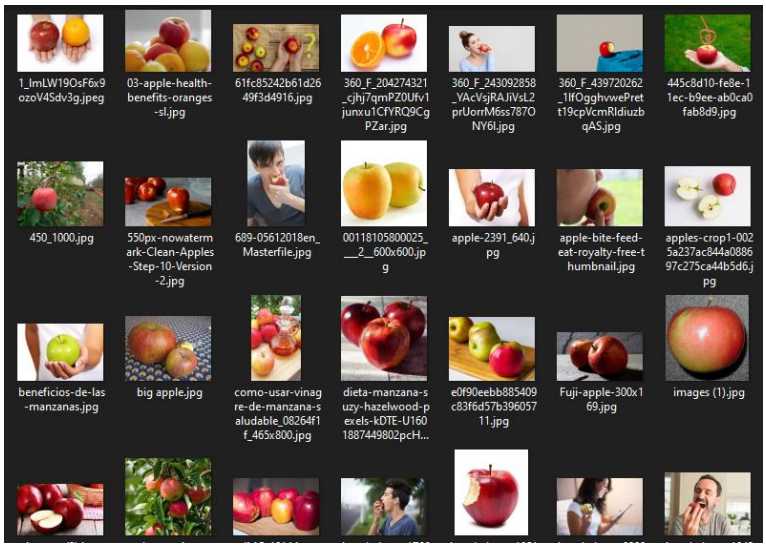


Figura 8. Set imágenes del objeto manzana

Crear dataset de entrenamiento

Es ejecutado mediante anotaciones que consisten en extraer metadatos de las imágenes los mismos que se obtienen de los puntos trazados para identificar el objeto dentro de la imagen.

El proceso en el interior de esta herramienta es fácil, demanda únicamente enmarcar un objeto del cual se requiere obtener la metadata y guardar el archivo XML que crea el sistema en formato Pascal VOC.

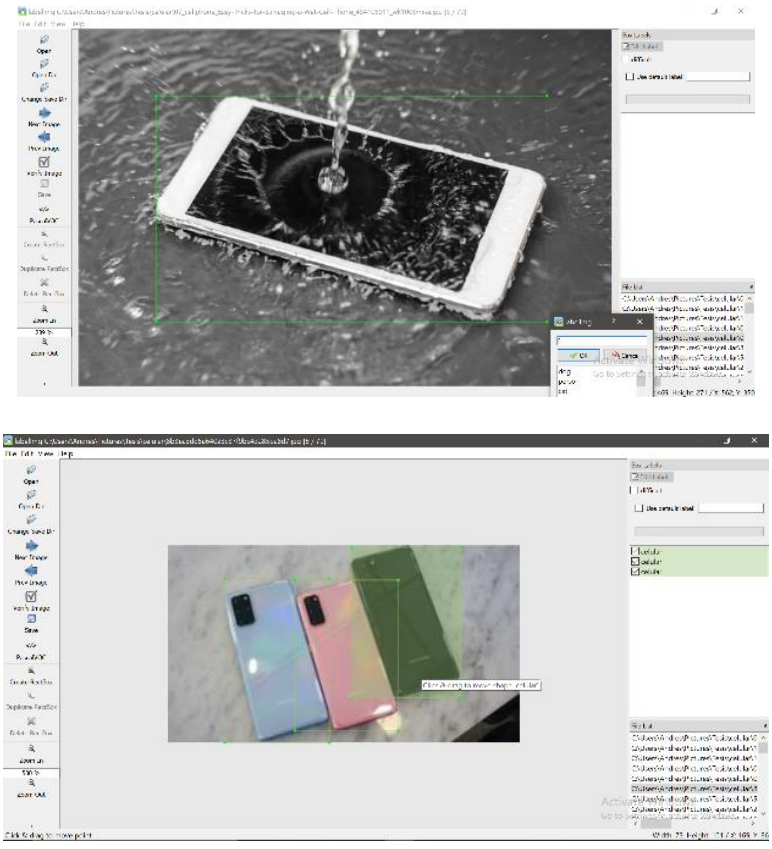


Figura 9. Software de anotación de imágenes LabelImg

Se obtiene un archivo XML como resultado del proceso, en formato Pascal VOC compuesto por los siguientes campos: carpeta, nombre del archivo, ubicación, tamaño, y características del objeto como la segmentación, nombre, posición, son difíciles de distinguir.

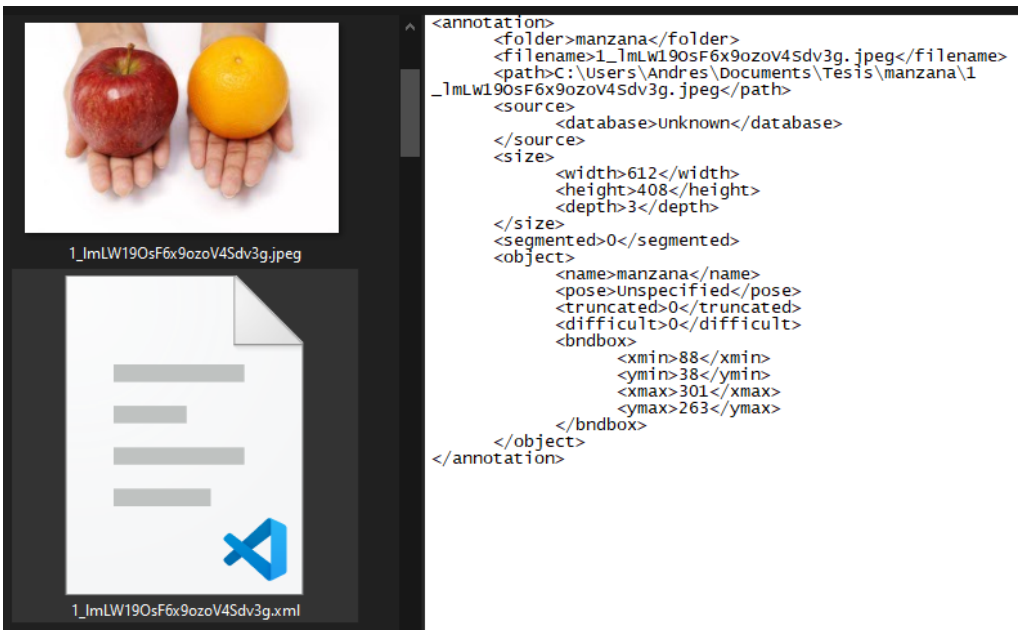


Figura 10. Archivo XML resultante del etiquetado de imágenes

Entrenamiento del modelo

Para abreviar el entrenamiento del modelo se utilizará la librería Tensorflow Lite Model Maker. Esta librería utiliza el aprendizaje de transferencia para reducir la cantidad de datos de entrenamiento requeridos y acortar la cantidad de tiempo de entrenamiento.

```
!pip install -q tflite-model-maker
```

```
!pip install -q tflite-support
```

Importar numpy para el tratamiento de datos numéricos en arreglos multidimensionales, e importar clases que permitan exportar el modelo, especificar el tipo de arquitectura y los parámetros el entrenamiento del modelo.

```
import numpy as np
import os

from tflite_model_maker.config import ExportFormat, QuantizationConfig
from tflite_model_maker import model_spec
from tflite_model_maker import object_detector

from tflite_support import metadata

import tensorflow as tf
assert tf.__version__.startswith('2')

tf.get_logger().setLevel('ERROR')
from absl import logging
logging.set_verbosity(logging.ERROR)
```

Figura 11. Librerías importadas para la ejecución del código

Las imágenes en train_data se utilizan para el entrenamiento del modelo de detección de objetos personalizado.

Mientras que, las imágenes en val_data se utilizan para comprobar si el modelo puede generalizar bien a imágenes que no ha visto antes.

```
train_data = object_detector.DataLoader.from_pascal_voc(
    'tesis/train',
    'tesis/train',
    ['celular', 'manzana', 'botella','taza', 'perro', 'gato', 'mujer', 'hombre']
)

val_data = object_detector.DataLoader.from_pascal_voc(
    'train/validate',
    'train/validate',
    ['celular', 'manzana', 'botella','taza', 'perro', 'gato', 'mujer', 'hombre']
)
```

Figura 12. Selección de elementos para entrenamiento y validación.

Selección de la arquitectura del modelo

EfficientDet-Lite[0-4] es una familia de modelos de detección de objetos aptos para dispositivos móviles/IoT derivados de la arquitectura EfficientDet .

Rendimiento de cada modelo en comparación con los demás:

Arquitectura modelo	Tamaño (MB)	Latencia (ms)	Precisión promedio
EfficientDet-Lite0	4.4	146	25.69%
EfficientDet-Lite1	5.8	259	30.55%
EfficientDet-Lite2	7.2	396	33.97%
EfficientDet-Lite3	11.4	716	37.70%
EfficientDet-Lite4	19.9	1886	41.96%

Tabla 13. Comparación de Arquitectura EfficentDet

Para este caso se utilizará EfficientDet-Lite0 para entrenar el modelo.

```
spec = model_spec.get('efficientdet_lite0')
```

Figura 13. Arquitectura seleccionada para el entrenamiento

Se establece epochs = 20, quiere decir que pasará por el conjunto de datos de entrenamiento 20 veces.

Se establece batch_size = 1 para que se recorran una a una las imágenes en el conjunto de datos de entrenamiento.

```
model = object_detector.create(train_data, model_spec=spec, batch_size=1, epochs=20, validation_data=val_data)
```

Figura 14. Parámetros establecidos para el entrenamiento

Evaluación del modelo

Después de entrenar el modelo de detección de objetos con las imágenes del grupo de datos de entrenamiento, se realizará la validación del conjunto de imágenes de prueba que el modelo no ha procesador previamente.

```
modelo.evaluar ( val_data )
```

Figura 15. Validación de datos

Exportar el modelo

Por último, se exporta el modelo de detección de objetos entrenado al formato TensorFlow Lite. Lo cual da paso a que el modelo TensorFlow Lite sea más liviano, se ejecute más rápido en la CPU Raspberry Pi

```
model.export(export_dir='.', tflite_filename='tesis.tflite')
```

Figura 16. Exportar el modelo entrenado

3.3 Implementación del prototipo de visión artificial basado en software libre y hardware de bajo costo

El prototipo se encuentra estructurado de la siguiente forma:

- Unas gafas que poseen una cámara ubicada en la parte frontal, ya que es de fácil uso, además brinda comodidad al usuario.
- Un computador Raspberry Pi.
- La batería que se conecta de forma fija a la Raspberry

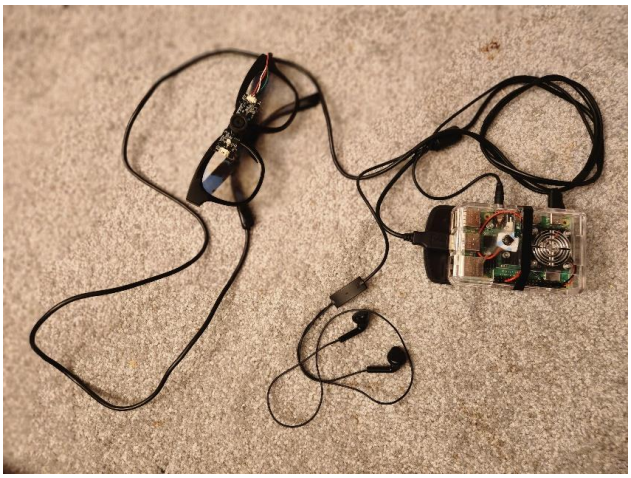


Figura 17. Prototipo de gafas de visión artificial.

3.3.1 Software del sistema

En lo que se refiere al software del prototipo el mismo se implementara en el sistema operativo nativo para raspberry Raspbian en su versión Buster, adicionalmente como lenguaje de programación se utilizara Python (3.6), el procesamiento de imágenes se procesara en base a las bibliotecas de OpenCV; entre otras, las librerías que se requieren para el desarrollo del software son RPi.GPIO, OpenCV, Tensorflow, Numpy, Thresholding, OS, etc. permiten el desarrollo de todo el sistema para obtener el reconocimiento de objetos y por ende la visión artificial.

Ejecución de archivos bash para iniciar el software.

La ejecución automática del programa se llevará a cabo haciendo uso del archivo autostart.desktop.

/home/pi/config/autostart/ □ file autostart.desktop

```
[Desktop Entry]
Exec=sudo bash /home/pi/runtesis.sh
```

En este archivo se programa la ejecución del archivo bash runtesis.sh

```
#!/bin/sh
clear
cd/home/pi/raspberry_pi/
python3 detect.py --model tesis.tflite
```

3.3.2 Configuración Raspberry PI

Se encuentra inicialmente instalado en una tarjeta MicroSD, de acuerdo a los requerimientos del proyecto se prevé trabajar con software libre; debido al tipo de procesamiento que se realizará es necesario modificar la configuración del swapfile, creando así una memoria virtual en el disco para la ejecución del código.

/etc/dphys-swapfile

Se establece la memoria virtual en 2048KB

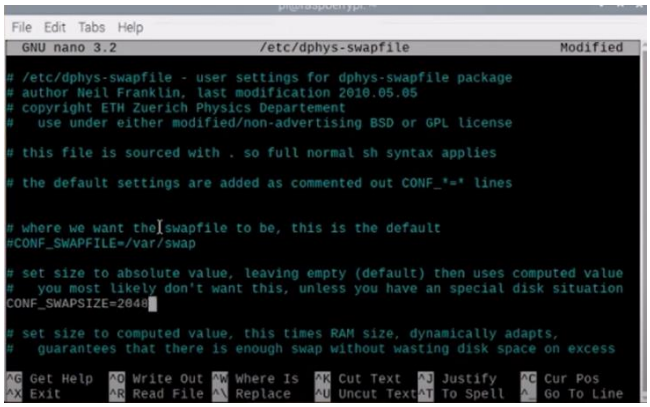
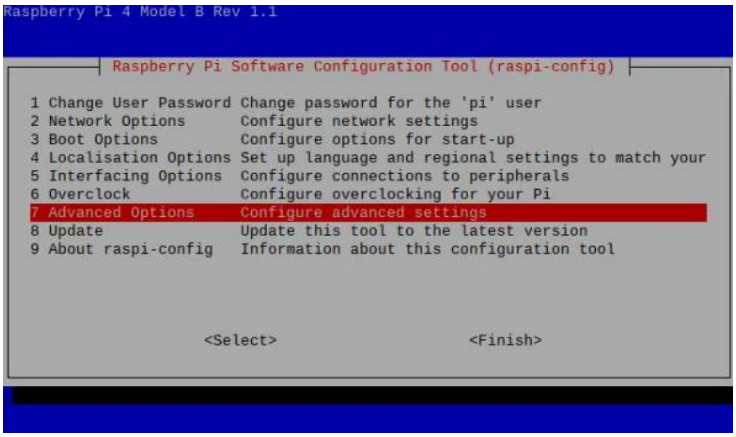


Figura 18. Modificar la memoria en el swapfile

Adicionalmente se verifica que todo el espacio de la RaspberryPI esté disponible para el uso.

sudo raspi-config



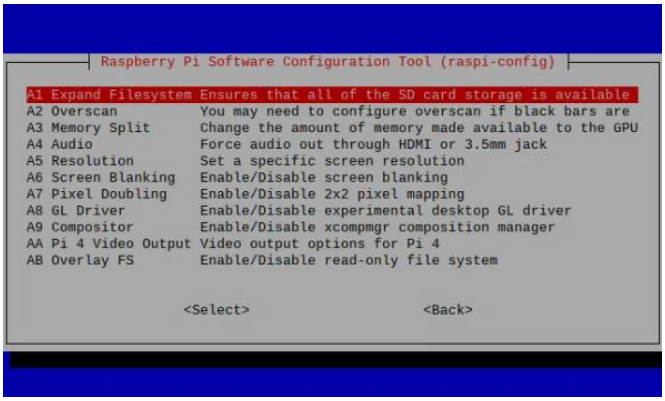


Figura 19. Configuración de Raspberry PI

Es necesario habilitar el uso de la cámara, para ello ingresamos a la configuración del sistema y se habilita el uso de la camara.

sudo raspi-config

Habilitar conexión de la cámara en Interface Options

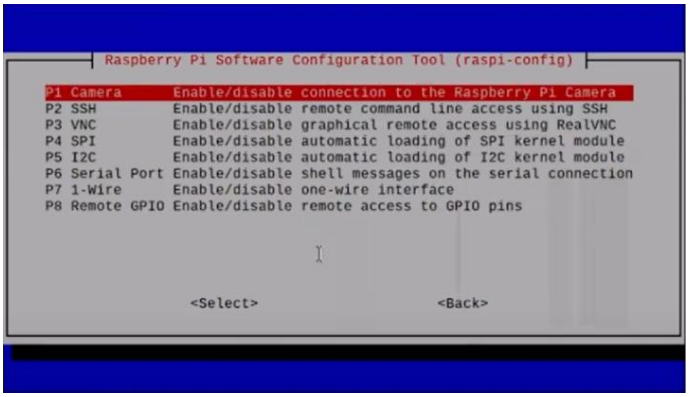


Figura 20. Habilitar conexión de la cámara

El procesamiento de imágenes se efectuará con el uso de OpenCV, esta librería es de código abierto, provee una infraestructura de visión artificial para el prototipo. Para facilitar la instalación de esta librería se realizara por medio de pip (Package Installer for Python).

sudo apt-get update && sudo apt-get upgrade

mkdir ~/src && cd ~/src

wget https://bootstrap.pypa.io/get-pip.py

$ sudo python3 get-pip.py

sudo pip install opencv-contrib-python

Para ejercer el reconocimiento de imágenes se utilizará las herramientas que se ofrecen en Tensorflow en su versión Lite al ser un proyecto que será desarrollado en un dispositivo móvil. Se procede a la instalación de esta librería con los siguientes comandos:

sudo apt-get update

sudo apt-get install python3-tflite-runtime

Es necesario instalar mediapipe y numpy

sudo pip3 install mediapipe.rpi4

Instalar librería de Python gtts (Google Text-to-Speech)

sudo pip3 install mediapipe.rpi4

Instalar reproductor de audio mpg123

sudo apt install mpg12

Interfaz de Usuario (R)

Consiste en un único botón de acción que, al presionarlo, automáticamente empieza a identificar objetos dentro de la captura de la imagen que se envía al procesamiento del sistema. La ubicación y diseño de este botón permite facilitar al máximo el uso del dispositivo.

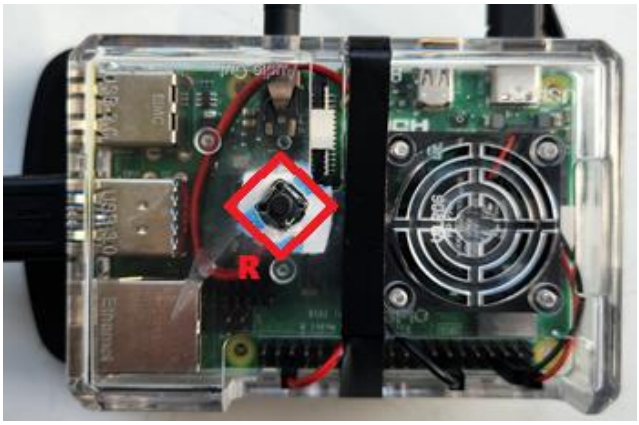


Figura 21. Interfaz de usuario, Botón R

Botón encendido / apagado (S)

La forma correcta de encender o apagar el dispositivo es como si se lo haría con cualquier otro computador de escritorio, ya que varias de las funciones que permiten el funcionamiento del equipo están relacionadas directamente al funcionamiento del sistema operativo. En la figura 22 a continuación, se muestra la ubicación del botón de encendido/apagado (S).



Figura 22. Botón encendido/apagado. Botón (S)

Entradas y Salidas a la Raspberry Pi.

La cámara con puerto USB, el cable de alimentación tipo “C” y el audífono conectado al puerto jack de 3.5 mm, deben permanecer interconectados para que el aparato trabaje de forma adecuada. Dichos puertos pueden ser removidos e intercambiados con otros periféricos.

A través de la Figura 22 se observa la salida de audio de la Raspberry Pi, el cable de alimentación tipo C, el puerto Jack, además se pueden reconocer la salida USB de la cámara.

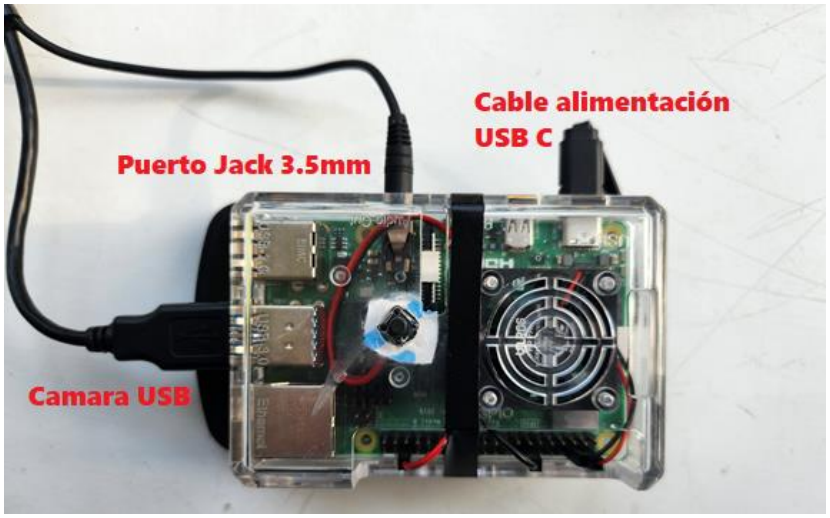


Figura 23. Entradas y salidas a la Raspberry Pi.

3.3.3 Batería Portatil 4000mAh (USB-C)

La batería tiene la función de alimentar a todo el sistema, del mismo modo se encarga de aportar energía a la Raspberry, así como a la cámara, audífonos y demás elementos.

A continuación, se detallan las características del sistema de energización adecuado para soportar los elementos que conforman el prototipo, estos puntos garantizan el correcto funcionamiento del equipo durante el tiempo descrito, ya que el procesador requiere un mínimo de corriente y voltaje para su funcionamiento, vale recalcar que el encendido y apagado del dispositivo se realiza por medio del botón incluido en la batería.

Sistema de energización

Mediante la Figura 24 se puede ver la fuente de alimentación del sistema. La cual es una batería de Li-Ion de 4000mA de capacidad, posee un puerto USB de carga.

Características

- Capacidad: 4500 mAh
- Corriente de salida de: 2.1 A
- Voltaje de salida: 12V
- Corriente de entrada: 1A
- Voltaje de entrada: 5V
- Tiempo de Duración: 2 horas

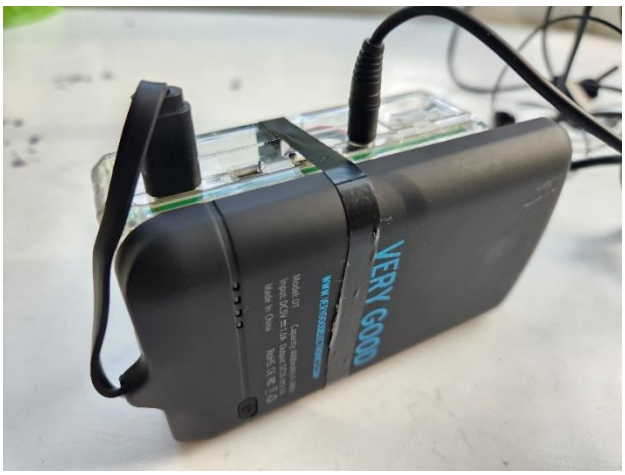


Figura 24. Batería 4000mAh, 5V 2.4A (USB-C)

3.3.4 Evaluación del prototipo

Iinicialmente, se efectuaron pruebas para verificar reconocimiento de objetos a través del sistema, tomando en cuenta los siguientes aspectos:

- Ambiente real, objetos aislados
- Porcentaje de detección
- Reconocimiento de objetos
- Tiempo de procesamiento

Tras el análisis del desempeño del dispositivo se procede a efectuar las pruebas a cerca del funcionamiento total del dispositivo, tomando como referencia a personas con discapacidad visual. Para extraer una evaluación del desempeño del prototipo y establecer las posibles mejoras en el prototipo.

Finalmente se introduce el análisis de los costos referenciales de la construcción del prototipo.

Pruebas con objetos aislados

El propósito de esta prueba fue conocer varios aspectos como: el porcentaje de detección, confiabilidad y tiempo que el sistema ocupa para la captura de un solo objeto. En función a los resultados se especificará la eficiencia del sistema.

Antes de llevar a cabo esta prueba se tomó en cuenta si el dispositivo pudo o no realizar un adecuado reconocimiento de los objetos, el porcentaje de acierto y cuánto tiempo se demoró el sistema procesar la información y emitir el audio al usuario. Además, se analizó las imágenes capturadas por el obturador, el cual sirvió para distinguir, tanto la etiqueta como la clase de

objeto y el porcentaje de confiabilidad con la que se estableció el reconocimiento del mismo.

Objeto de prueba	Imagen Referencial	Reconocimiento SI - NO	Porcentaje de acierto	Tiempo
Botella		SI	56%	1.27 s
Celular		SI	39%	1.11 s
Taza		SI	67%	1.09 s
Manzana		SI	73%	1.17 s

Tabla 14. Evaluación del prototipo en objetos reales

En la Tabla 14 se muestran las respuestas acerca de la respuesta promedio de reconocimiento de objetos, adquiriendo los resultados en porcentaje de reconocimiento, confiabilidad de acuerdo a la detección, así como el tiempo de procesamiento en cada prueba realizada con los objetos aislados.

Respuesta promedio Reconocimiento de objetos	
Reconocimiento	100%
Confiabilidad	58.75%
Tiempo	1.16 s

Tabla 15. Respuesta promedio detección

A través de los resultados expuestos en la Tabla 15 se determina que, el prototipo posee un buen desempeño, puesto que el porcentaje de reconocimiento fue del 100%, aparte de que el promedio de la confiabilidad obtenido fue del 58.75 %.

3.3.5 Pruebas en personas con discapacidad visual

Con el objetivo de conocer la respuesta que tiene el prototipo en un ambiente real se procede a realizar pruebas con personas que poseen algún nivel de discapacidad visual, durante las pruebas se ubicaron varios objetos, en posiciones diferentes y ángulos de visión que permitan evaluar el dispositivo en situaciones reales.



Figura 25. Foto prueba del dispositivo con personas

En la Figura 25 se puede observar una de las pruebas aplicadas a un hombre de 59 años, a quien se le sometió a evaluación varios tipos de objetos y personas para comprobar la eficacia del dispositivo. Finalizando con resultados satisfactorios que se muestran en la Tabla 16 donde se revelan los resultados conseguidos del sistema de reconocimiento practicados a dos personas con discapacidad visual.

	Objeto	SI	NO
Persona 1	Botella	X	
	Celular	X	
	Taza	X	
	Libro		X
Persona 2	Mujer	X	
	Perro		X
	Botella	X	
	Manzana	X	

Tabla 16. Resultados pruebas aplicadas a personas

Las respuestas registradas en la Tabla 16, presentan resultados de detección de objetos. Según estas cifras: en su mayoría los objetos fueron detectados de forma correcta.

Las personas estuvieron en un ambiente real lo que permitió evaluar la captura de las imágenes, considerando las variaciones de luz, debido a los diferentes ángulos de visión.

No obstante, los porcentajes de objetos acertados resultó elevado, razón por la cual se puede determinar que el prototipo funciona de forma correcta al considerar que se muestra como primera versión y las características del hardware se utilizaron de forma óptima, en relación a la calidad y el precio.

3.3.6 Costo referencial del prototipo

Para finalizar esta fase, en la Tabla 17, se exponen los costos referenciales de los elementos utilizado para la implementación del sistema.

Elemento	Precio
Raspberry PI4	$120.00
Cámara USB	$43.00
Tarjeta MicroSD 64GB	$18.00
Batería Inalámbrica 4800mAh	$25.00
Audífonos	$5.00
Pulsador	$1.00
Total	$212.00

Tabla 17. Costo referencial del prototipo

Capítulo 4: Conclusiones y Recomendaciones

En el último capítulo se demuestra las conclusiones y se propone algunas recomendaciones que pueden servir como base para futuros estudios que correspondan al presente tema de titulación. Tanto conclusiones como recomendaciones hacen hincapié en el diseño, implementación, pruebas del sistema de reconocimiento de objetos y el hardware de todo el prototipo.

4.1 Conclusiones

- Los continuos acercamientos y socializaciones con los usuarios permitieron identificar de forma clara las características y los objetivos de detección, para la elaboración del prototipo final, permitiendo la elaboración de un producto accesible que cuenta con amplias las funcionalidades descritas a lo largo del presente proyecto y de cierta forma personalizado hacia los potenciales usuarios de este dispositivo.

- Raspberry Pi 4B, en la relación costo y beneficio demostró ser un ordenador optimo, y resultó ser el más conveniente, ya que actualmente es uno de los más veloces y cuenta con características en hardware que benefician al prototipo.

- Acerca de la batería es necesario mantener el uso del dispositivo en base a las características propuestas en el documento ya que de lo contrario el prototipo se podría ver afectada de forma negativa al ser un computador necesita estabilidad en el flujo de corriente, así como de voltaje.

- Se pudo inferir que el hecho de usar librerías como: OpenCV y Tensorflow, sirven de aporte para programar el sistema de visión artificial puesto que, cuentan con una amplia variedad de herramientas para facilitar la programación de este tipo de proyectos y además su uso no implican gastos adicionales ya al ser software libre su uso no implica un pago.

- Es preciso que, en el proceso del entrenamiento de detección de objetos, la base de datos disponga no solo de una gran cantidad de imágenes, sino que éstas deben una tener buena calidad, relacionando proporcionalmente al tamaño de la imagen, ya que este proceso implica una mayor fiabilidad durante el reconocimiento de objetos.

- Al término del presente trabajo de investigación, se concluye que se pudo cumplir satisfactoriamente con todos los objetivos y alcances proyectados.

4.2 Recomendaciones

- Si se desea desarrollar un proyecto, es importante realizar un análisis de los requerimientos y la metodología que se aplicara al mismo, ya que de esta manera se puede realizar una planificación adecuada y evitar errores durante el desarrollo que requieran retomar o reconsiderar cambios en el proyecto.
- En cuanto al dispositivo se sugiere que jamás se desconecte directamente la energía, así se podrán evitar daños en la tarjeta micro SD, la cual contiene todos los programas colocados en el sistema operativo.
- Al momento de implementar el tipo de funcionalidades que posee este proyecto, así como otros basados en nuevas áreas de tecnología se recomienda hacer uso de librerías que cuenten con documentación amplia y de fácil acceso, para evitar inconvenientes durante el desarrollo del mismo.
- Se recomienda que la universidad incentive a sus alumnos a participar en proyectos multidisciplinarios, que involucren la aplicación de nuevas tecnologías a otro tipo de áreas que pueden verse beneficiadas y hacer uso de forma accesible a estos desarrollos.

Bibliografía:

Aviolita, E (2019). Pendeteski Penyit Tanaman Buha Kamera Berbasis Rasberry Pi. (Tesis doctoral, Instituto de Tecnología de Malan- Indonesia.

Clarke, R. J. (1985) Transform Coding of Images. Academic Press.R. González y R. E. Woods, Digital Image Processing, Prentice Hall. 2002.

Digón A, Ferrer P. (2014) Configuración y ajustes de Sistemas de Sonido. Ediciones de la U.

F. G. Becker et al. (2015). Técnicas y Algoritmos Básicos de Visión Artificial, vol. 7, no. 1. https://www.researchgate.net/publication/269107473_What_is_governance /Civil wars_12December2010.pdf%0Ahttps://thinkasia.org/handle/11540/8282%0Ahttps://www.jstor.org/stable/41857625

G. Mariela et al. (2021). "Redes neuronales paralelas aplicadas a la visión computacional," vol. 273, pp. 725–728. http://sedici.unlp.edu.ar/handle/10915/12040

Gardner, Howard (1999). "Intelligence Reframed: Multiple Intelligences for the 21st Centur". Basic Books 2000.

Gutiérrez, J. (2014).*¿Qué es un framework web?*

http://www.lsi.us.es/~javierj/investigacion_ficheros/Framework.pdf

Hartnell T, (1986) Inteligencia Artificial, Procedimientos y Programas. Anaya Multimedia.

Herrera F, (2014) Inteligencia Artificial, Inteligencia Computacional y Big Data, Universidad de Jaén.

Huang E, (1975) Picture Processing and Digital Filtering, Springer-New York.

Hunt B, Cannon T, (1976) Non-Stationary Assumptions for Gaussian Models of Image, IEE.

IDEAUNO. (2014). *Desarrollo de Prototipos*. IDEAUNO. https://www.ideauno.cl/desarrollo-de-prototipos-web

López D, (2009). Ingeniería del Sonido. Starbook.

López, R. (2015). Metodologías Ágiles de Desarrollo de Software. *Revista Tecnológica.* https://core.ac.uk/download/pdf/80296686.pdf

Maida, E., & Pacienzia, J. (2015). *Metodologías de desarrollo de software* [Pontificia Universidad Católica Argentina]. https://repositorio.uca.edu.ar/bitstream/metodologias-desarrollo- software.pdf

Quintana, M. (2015) Discapacidad Visual, sugerencias socioeducativas. Trillas.

Rodríguez R, Sossa J. (2012) Procesamiento y Análisis Digital de Imágenes. Alfaomega.

Software Engineering Institute. (2021, May 10). *Software Architecture*. Carnegie Mellon University. https://www.sei.cmu.edu/our-work/software-architecture/

Sommerville, I. (2005). *Ingeniería del software* (M. Romo (ed.); 7ª ed). Pearson.

Sossa J. (2013) Visión Artificial. Rasgos Descriptores para el Reconocimiento de Objetos. Ra-Ma.

Tecnologías Información. (2019). *Gestión de datos: Conceptos, Mejores Prácticas y Beneficios*. Tecnologías Información. https://www.tecnologias-informacion.com/gestiondatos.html

Anexos

MANUAL DE USUARIO

PROPÓSITO

El presente manual tiene la finalidad de ser una guía de operación del prototipo, permitiendo al lector del mismo obtener las destrezas y conocimientos necesarios para su funcionamiento, y ser una herramienta de consulta a la cual el usuario pueda recurrir en cualquier momento.

CARACTERÍSTICAS

Las partes constitutivas del prototipo se dividen en dos secciones que son el hardware y el software, a continuación, se detallan casa una de ellas:

Hardware

- Raspberry PI 4 Model B: Es un miniordenador sin todo el hardware que compone a un computador de escritorio y se caracteriza por su simplicidad, su flexibilidad, la multitud de accesorios compatibles y su bajo coste. El modelo 4 B es el modelo más moderno, y también el más potente y versátil. Tiene conector Gigabit Ethernet y una tarjeta WiFi + Bluetooth integrada; conector de alimentación del tipo USB-C y dos conectores MIPI; conexión GPIO a la que se pueden conectar una gran variedad de dispositivos externos, en este caso el botón de reconocimiento de objetos.

- Cámara USB: El punto de acceso de información con el exterior, es la cámara cuyo propósito principal es capturar imágenes una vez que el usuario presione el botón S. La cámara USB seleccionada cuenta con video Full HD de 1080p que permite capturar imágenes con resolución de 15 Mpx y una tasa de fotogramas de 30 cuadros por segundo.

- Audífonos: Los audífonos no requieren características específicas ya la Raspberry permiten conexiones directas al puerto Jack 3.5 o por medio de dispositivos USB.

- Batería: La batería del sistema cuenta con capacidad de 4000mAh, compuesta de Ion Litio, posee un puerto USB de carga y un puerto USB tipo C como salida de la fuente de alimentación.

Software

- Raspbian Buster: SO lanzado como sucesor de Raspbian Stretch, que fue el sistema operativo predeterminado de Raspberry Pi durante casi 2 años, es compatible con todos los modelos de Raspberry Pi. Buster se basa en la última versión de Debian Linux 10,

con kernel de Linux versión 4.19 y compilador GCC 8.3. Al igual que los sistemas operativos Raspbian anteriores.

- Python: Python es un lenguaje de programación de propósito general y alto nivel. Su filosofía de diseño enfatiza la legibilidad del código con el uso de sangría significativa. Python se tipifica dinámicamente y se recolecta basura. Admite múltiples paradigmas de programación, incluida la programación estructurada, orientada a objetos y funcional.
- Tensorflow Lite: Es una biblioteca para implementar modelos en dispositivos móviles, microcontroladores y otros dispositivos perimetrales. Esta biblioteca nos permite implementar modelos de aprendizaje automático en este caso en particular de visión artificial.

MODO DE USO

Con el fin de facilitar al máximo el uso de este prototipo, únicamente el usuario cuenta con dos botones de acción y un puerto de entrada para cargar la batería.

En cuanto a la forma de uso de la cámara, la misma esta acoplada sobre el soporte de unas gafas de uso común y a esta se encuentran unidos los audífonos.

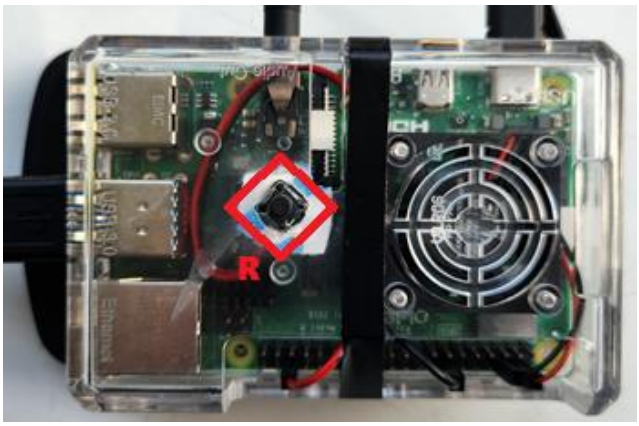
Botón Encendido – Apagado (S)

El dispositivo cuenta con un botón (S) para el encendido - apagado del mismo, este se encuentra ubicado en la parte posterior de la batería, como se muestra en la siguiente imagen.



Botón Reconocimiento (R)

El botón (R) de reconocimiento de objetos se encuentra situado sobre la carcasa de la Raspberry, al accionar este botón se envía la señal para obtener una imagen real del video continúo transmitido por la cámara, para que sea procesada por los algoritmos de reconocimiento de objetos y posteriormente se reproduzca el nombre y la cantidad de objetos reconocidos al oído del usuario.



Encuesta – Identificación de objetivos y características del prototipo

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR
FACULTAD DE INGENIERÍA
ESCUELA DE SISTEMAS Y COMPUTACIÓN

DISEÑO E IMPLEMENTACIÓN DE UN PROTOTIPO BASADO EN INTELIGENCIA ARTIFICIAL PARA EL RECONOCIMIENTO DE OBJETOS EN EL ENTORNO DE UNA PERSONA NO VIDENTE

ENCUESTA

Edad: _____

1. Como le gustaría que fuese un dispositivo de visión artificial:
   OPCIÓN ÚNICA
   a) Gafas de uso común _____ ( )
   b) Pulsera con cámara _____ ( )
   c) Dispositivo colgante en el pecho del usuario _____ ( )
   d) Aplicación disponible desde el celular _____ ( )

2. ¿En cuanto al precio, cual es el rango que estaría dispuesto a pagar?
   OPCIÓN ÚNICA
   a) $50 - $100 _____ ( )
   b) $100 - $200 _____ ( )
   c) $200 – mas _____ ( )

3. ¿Cómo prefiere que sea su modo de uso?
   OPCIÓN ÚNICA
   a) Botón pulsador _____ ( )
   b) Varios botones de acciones _____ ( )
   c) Acción por voz _____ ( )

4. ¿Cuáles son las características que considera más importantes de acuerdo a la batería?
   OPCIÓN MULTIPLE
   a) Dispositivo autónomo _____ ( )
   b) Baterías recargables _____ ( )
   c) Baterías de larga duración _____ ( )
   d) Costo accesible _____ ( )
   e) Tiempo de carga _____ ( )

5. ¿Cuál es la relación que considera más importante para el dispositivo?
   OPCIÓN ÚNICA
   a) Alta velocidad de respuesta – Baja precisión de reconocimiento _____ ( )
   b) Baja velocidad de respuesta – Alta precisión de reconocimiento _____ ( )

6. Seleccione las características considera más importantes para el dispositivo
OPCIÓN MULTIPLE
   a) Autónomas _____ ( )
   b) Recargable _____ ( )
   c) Interfaz de usuario sencilla _____ ( )
   d) Fácil transportación _____ ( )
   e) Alta precisión _____ ( )
   f) Fácil mantenimiento _____ ( )
   g) Económicas _____ ( )
   h) Otros: (Detalle las características) _____
   _____
   _____
   _____
   _____
   _____

7. Cuáles de las siguientes considera las categorías más importantes para el reconocimiento de objetos.
OPCIÓN MULTIPLE
   a) Medicinas _____( )
   b) Animales _____ ( )
   c) Dinero _____ ( )
   d) Personas _____ ( )
   e) Dispositivos electrónicos _____( )
   f) Alimentos _____( )
   g) Utensilios _____( )
   h) Maquillajes _____ ( )
   i) Muebles _____ ( )
   j) Objetos de aseo personal _____( )

8. Cuáles de los siguientes objetos considera necesarios para el reconocimiento de objetos.
OPCIÓN MULTIPLE

TV		Silla	
Manija		Jabón de baño	
Celular		Botella	
Frutas y vegetales		Billetes	
Laptop		Puerta	
Monedas		Cubiertos	
Cama		Perro	
Mochila		Personas	
Zapatos		Ropa	
Baston		Esfero	