



Desarrollo de un prototipo que brinde accesibilidad a la composición de una imagen digital preprocesada mediante el seguimiento de la posición dactilar para personas con discapacidad visual

Curicho Oyos, Lilian Gisela

Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Trabajo de titulación, previo a la obtención del título de Ingeniera en Electrónica y Telecomunicaciones

Ing, Larco Bravo, Julio César, MSc.

14 de agosto de 2023

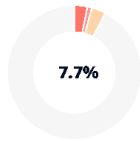


04_Escrito_Lilian Curicho_2.pdf

Scan details

Scan time: August 14th, 2023 at 13:54 UTC Total Pages: 55 Total Words: 13504

Plagiarism Detection



Types of plagiarism		Words
● Identical	3.1%	421
● Minor Changes	1.1%	149
● Paraphrased	3.5%	466
● Omitted Words	0%	0

AI Content Detection



Text coverage
● AI text
○ Human text

🔍 Plagiarism Results: (100)

🌐 **Preguntas frecuentes – CEISH – Ministerio de Salud ...** 1.1%
<https://www.salud.gob.ec/preguntas-frecuentes-ceish/>
 Ministerio de Salud Pública ...

📄 **(PDF) Muñoz Chuquimarca, Andy Gustavo Departa...** 1%
<https://dokumen.tips/documents/muoz-chuquimarca-andy-g...>
 Others
 ...

🌐 **Definiciones | Racionalidad Ltda** 0.8%
<https://racionalidadltda.wordpress.com/2023/03/19/definicio...>
 Racionalidad Ltda "Aquí no reina la codicia, la gula ni la avaricia. El dinero no es caliente, todo es transpar..."





Departamento de Eléctrica, Electrónica y Telecomunicaciones
Carrera de Ingeniería en Electrónica y Telecomunicaciones
Certificación

Certifico que el trabajo de titulación: **“Desarrollo de un prototipo que brinde accesibilidad a la composición de una imagen digital preprocesada mediante el seguimiento de la posición dactilar para personas con discapacidad visual”** fue realizado por la señorita **Curicho Oyos, Lilian Gisela**; el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 14 de agosto de 2023



Ing. Larco Bravo, Julio Cesar, MSc.

C.C.:1710638808



Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Responsabilidad de Autoría

Yo, **Curicho Oyos, Lilian Gisela**, con cédula de ciudadanía n° 0503482895, declaro que el contenido, ideas y criterios del trabajo de titulación: **“Desarrollo de un prototipo que brinde accesibilidad a la composición de una imagen digital preprocesada mediante el seguimiento de la posición dactilar para personas con discapacidad visual”** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 14 de agosto de 2023

Lilian Curicho

Curicho Oyos, Lilian Gisela

C.C.: 0503482895



Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Autorización de Publicación

Yo, **Curicho Oyos, Lilian Gisela**, con cédula de ciudadanía n° 0503482895, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **“Desarrollo de un prototipo que brinde accesibilidad a la composición de una imagen digital preprocesada mediante el seguimiento de la posición dactilar para personas con discapacidad visual”** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 14 de agosto de 2023

Lilian Curicho

Curicho Oyos, Lilian Gisela

C.C.: 0503482895

Dedicatoria

Este trabajo de titulación sin duda es dedicado a mi madre Alicia y a mi padre Alfredo, que hasta el último día de este proceso han permanecido sosteniéndome con su amor.

Lilian Gisela Curicho Oyos

Agradecimiento

Agradezco a un ser especial Dios que, por su presencia constante y amor incondicional mi corazón hoy en día está lleno de emoción por este logro académico y que junto a mis padres Alicia y Alfredo me han infundido fe, fortaleza y perseverancia para cumplir cada una de mis metas. Doy un profundo agradecimiento a mi hermana Nataly, mi hermano Diego y mi sobrina Yuli que siempre han estado junto a mí al igual que mis padres, consintiéndome y dándome vibras positivas para llegar a este momento.

De igual manera agradezco a Ing. Julio Larco por permitirme desarrollar este trabajo de titulación junto a él y aportarme su sabiduría y gran conocimiento.

Y finalmente quiero expresar mi gratitud a mis amigos que han permanecido a lo largo del tiempo con su apoyo inquebrantable siendo mis compañeros de clase y de aventuras.

Lilian Gisela Curicho Oyos

Índice de contenidos

Índice de tablas	12
Índice de figuras	13
Resumen	15
Abstract	16
Capítulo I	17
Descripción	17
Introducción	17
Justificación e importancia.....	17
Estado del arte	20
Alcance.....	21
Objetivos	22
Objetivos generales.....	22
Objetivos específicos	22
Organización del trabajo	23
Capítulo II	24
Marco teórico	24
Discapacidad visual	24
Procesamiento digital de imágenes.....	25

Orientación	26
Redimensión.....	27
Transformación de perspectiva.....	27
Inteligencia Artificial	28
Machine Learning	29
Deep Learning	30
Artificial Vision	31
Detección de la posición dactilar	32
Python.....	33
Raspberry Pi.....	36
Raspberry Pi 4 Model B	36
Cámara.....	38
Capítulo III	40
Metodología de desarrollo del proyecto	40
Configuración de la Raspberry Pi 4 Model B	41
Bibliotecas.....	42
Diagrama general del aplicativo.....	42
Ingreso de imagen preprocesada.....	45
Preprocesamiento.....	46
Importación de imagen	48

	10
Etiquetado de las zonas de interés de la imagen	49
Conexión de puntos	50
Definición del etiquetado	51
Eliminación del último punto	52
Almacenamiento de información en una base de datos	53
Generación de código QR.....	54
Descarga de imagen en formato específico	54
Corrección de orientación de la imagen	55
Carga de información asociada a la imagen.....	57
Seguimiento de la posición dactilar	57
Detección de la posición dactilar en la posición de interés.....	57
Reproducción de mensaje asociado a la región de interés	58
Capítulo IV	59
Pruebas y resultados	59
Metodología	59
Participantes.....	60
Procedimiento de pruebas.....	60
Fase preliminar	62
Fase total.....	64
Capítulo V	73

Conclusiones y trabajos fututos	73
Conclusiones	73
Trabajos futuros	74
Bibliografía	76

Índice de tablas

Tabla 1 <i>Especificaciones técnicas de la Raspberry Pi 4 Model B</i>	37
Tabla 2 <i>Especificaciones técnicas de la cámara</i>	39
Tabla 3 <i>Fase preliminar - Registro de zona de interés identificadas</i>	62
Tabla 4 <i>Fase total - Registro de zona de interés identificadas</i>	65
Tabla 5 <i>Registro de puntuación de la encuesta para la aceptación del prototipo</i>	70

Índice de figuras

Figura 1 <i>Ilustración de la estructura del prototipo para la descripción de la composición de una imagen digital preprocesada</i>	19
Figura 2 <i>Una imagen preprocesada con una zona de interés</i>	20
Figura 3 <i>Coordenadas espaciales</i>	26
Figura 4 <i>Transformación de perspectiva</i>	28
Figura 5 <i>Campos dentro de la Inteligencia Artificial</i>	29
Figura 6 <i>Microcomputadora Raspberry Pi 4 Model B</i>	37
Figura 7 <i>Cámara web</i>	38
Figura 8 <i>Información del Sistema Operativo de la Raspberry Pi 4</i>	41
Figura 9 <i>Diagrama de bloques de la aplicación de la descripción auditiva de la zona de interés según la posición</i>	43
Figura 10 <i>Diagrama de flujo del aplicativo de la descripción auditiva de la zona de interés según la posición dactilar sobre una imagen</i>	44
Figura 11 <i>Ingreso de imagen preprocesada</i>	45
Figura 12 <i>Diagrama de bloques del preprocesamiento de una imagen</i>	47
Figura 13 <i>Interfaz de la aplicación para el preprocesamiento de la imagen</i>	48
Figura 14 <i>Importación de imagen a preprocesar</i>	49
Figura 15 <i>Selección de puntos en la imagen de la zona de interés</i>	50
Figura 16 <i>Formación del polígono de la zona de interés</i>	50
Figura 17 <i>Etiquetado de la zona de interés</i>	51
Figura 18 <i>Ausencia de la etiqueta o zona de interés para el almacenamiento</i>	52
Figura 19 <i>Almacenamiento de datos de etiquetado de la imagen</i>	53
Figura 20 <i>Identificador para acceder a la base de datos almacenado</i>	54

Figura 21 <i>Material generado para la persona con discapacidad visual</i>	55
Figura 22 <i>Puntos céntricos de los marcadores de ArUco para estimar la orientación</i>	56
Figura 23 <i>Transformación de perspectiva en tiempo real</i>	56
Figura 24 <i>Nombres de los 21 puntos de referencia de una mano</i>	57
Figura 25 <i>Posición del dedo índice en la zona de interés en tiempo real</i>	58
Figura 26 <i>Evaluación de validación del prototipo con personas con discapacidad visual</i>	61
Figura 27 <i>Zonas de interés válidas en la zona de interés correspondientes en la fase preliminar</i>	63
Figura 28 <i>Concordancia de coordenadas entre lo seleccionado y en tiempo real</i>	64
Figura 29 <i>Zonas de interés válidas en la zona de interés correspondientes en la fase última y definitiva</i>	68
Figura 30 <i>Formulación inicial de TAM</i>	69

Resumen

La mayoría de información que se percibe a diario es de manera visual en forma de imágenes que, a través de esta, se puede ilustrar ideas y conceptos e incluso plasmarse en la mente. Las personas no videntes necesitan de un asistente vidente que proporcione información de la imagen para que logren una visualización mental de la misma según su composición, es por ello que la prioridad es brindar soluciones inclusivas con nuevas tecnologías para mitigar la discriminación, las brechas en el aprendizaje y fomentar la inclusión digital para promover y garantizar la autoeducación de estas personas con la disposición de un sistema que permita la interacción del no vidente con una imagen preprocesada y evitar del asistente.

El presente proyecto tiene como objetivo la descripción de la composición de una imagen digital preprocesada relacionada al seguimiento de la posición dactilar en tiempo real para personas con discapacidad visual, usando herramientas de Inteligencia Artificial (IA) ejecutado en una microcomputadora.

En este proyecto se desarrolló un prototipo que consta de componentes físicos y sistemas lógicos llamados hardware y software, donde el hardware estará compuesto de un microcomputador, un altavoz y una cámara de video, mientras que el software constará de la tecnología del aprendizaje automático perteneciente a la rama de la Inteligencia Artificial (IA), La IA ayuda a la ejecución del desarrollo del software que consiste en la detección de la posición dactilar sobre una imagen preprocesada, la cual tiene etiquetas previamente colocadas sobre la zona de interés de composición. Cuando la posición dactilar se encuentre sobre una zona de interés se reproduce un audio que describa la misma, la cual brinda accesibilidad a su composición con el propósito de crear una experiencia de poseer un asistente que describa la imagen guiada por la detección del dedo índice.

Palabras Clave: Imagen, posición dactilar, zona de interés, inteligencia artificial, text to speech

Abstract

Most of the information that is perceived on a daily basis is visually in the form of images that, through this, ideas and concepts can be illustrated and even captured in the mind. Blind people need a sighted assistant who provides information about the image to achieve a mental visualization of it according to its composition, which is why the priority is to provide inclusive solutions with new technologies to mitigate discrimination, learning gaps and encourage digital inclusion to promote and guarantee the self-education of these people with the provision of a system that allows the interaction of the blind with a pre-processed image and avoids the assistant.

The objective of this project is to describe the composition of a preprocessed digital image related to real-time fingerprint position tracking for people with visual disabilities, using Artificial Intelligence (AI) tools executed on a microcomputer.

In this project, a prototype will be developed that will consist of physical components and logical systems called hardware and software, where the hardware will be composed of a microcomputer, a speaker and a video camera, while the software will consist of machine learning technology belonging to In the branch of Artificial Intelligence (AI), AI helps to execute the development of the software that consists in the detection of the fingerprint position on a preprocessed image, which has labels previously placed on the area of composition interest. When the finger position is over an area of interest, an audio describing it is played, which provides accessibility to its composition with the purpose of creating an experience of having an assistant that describes the image guided by the detection of the index finger.

Keywords: Image, finger position, area of interest, artificial intelligence, text to speech

Capítulo I

Descripción

Introducción

En el mundo la humanidad posee diversas capacidades, habilidades y desafíos únicos independientemente de su condición física e intelectual, sin embargo, la sociedad aún carece de empatía hacia las personas que tienen capacidades diferentes sin imaginarse que son seres humanos con sus propias fortalezas tal cómo lo tienen todos y por ende hay que priorizar y garantizar que absolutamente todas las personas, independientemente de su capacidad, gocen de las mismas oportunidades para el desarrollo de su potencial.

La tecnología es parte del desarrollo de oportunidades porque ayuda a superar barreras y ejercer una vida digna, aportando inclusión con las herramientas que estas pueden proporcionar a las personas que tienen capacidades diferentes cómo es el caso de la discapacidad visual, permitiendo la accesibilidad a ayudas visuales según el grado de discapacidad, por tanto existe ayudas ópticas, electro ópticas, instrumentos de lectura, acceso a la información y movilidad personal (Nieto, 2012). Cada una de las ayudas mencionadas van aumentando su tecnología, tanto que, en la actualidad éstas tienen avances en la inteligencia artificial para automatizar procesos y permitir la accesibilidad a dispositivos y dar una percepción más real.

La educación es el único distintivo que marca la diferencia al proporcionar conocimientos y habilidades a las personas con capacidades diferentes, optando por la oportunidad de romper aquella exclusión social y discriminatoria.

Justificación e importancia

La educación es indispensable para cumplir con los deberes del Estado Ecuatoriano y no importa la condición del ser humano para tener aquel derecho. La discapacidad visual en el

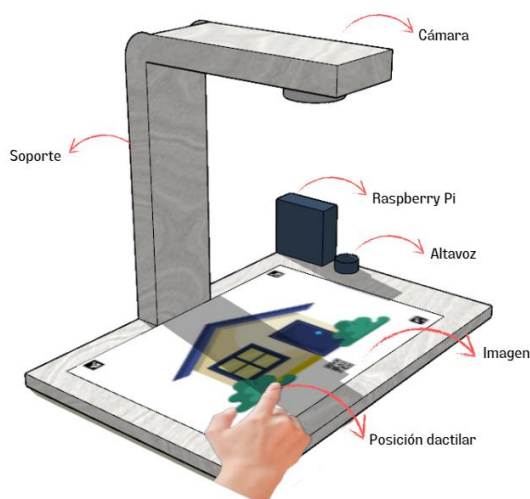
Ecuador representa un limitante por falta de aplicaciones tecnológicas que ayuden a acceder a la información para tener un aprendizaje adecuado creando un obstáculo que no permiten el desarrollo adecuado de las competencias de la población con discapacidad visual.

En el Ecuador de acuerdo con el Consejo Nacional para la Igualdad de Discapacidades (CONADIS) existen 471.205 personas con algún tipo de discapacidad, de este total un 19,75% presenta un nivel leve y severo, incluso el 11,54% presente un tipo de discapacidad visual (CONADIS, 2022). Esto representa un grupo de personas que tienen el derecho a recibir educación y capacitación que aporte a su desarrollo personal. Esto se logra gracias a que la tecnología hoy en día puede apoyar a la reducción de la brecha digital entre personas videntes y no videntes y apoyar a mejorar su inclusión en la sociedad a través de la educación.

Este proyecto tiene como fin ayudar a personas con discapacidad visual a través de un prototipo que brinde accesibilidad la composición de una imagen digital preprocesada según la ubicación de la posición dactilar en tiempo real y así el usuario pueda acceder a información sin ayuda de un tutor vidente como se ilustra en la Figura 1.

Figura 1

Ilustración de la estructura del prototipo para la descripción de la composición de una imagen digital preprocesada

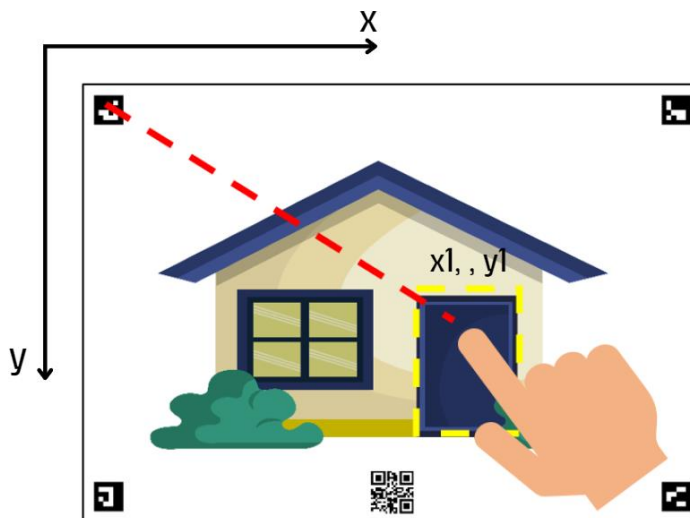


Este proyecto identifica la posición dactilar sobre una imagen preprocesada utilizando técnicas de Machine Learning, puntualmente de algoritmos de Deep Learning que gracias a un conjunto de datos de entrenamiento permiten controlar el proceso de aprendizaje y seleccionar el mejor modelo (Jácome & Viracocha, 2019). La imagen preprocesada se obtiene por la marcación de las zonas de interés dentro de la imagen a la cual se le asigna una etiqueta. Esta etiqueta, asociada a una zona de la imagen en unión a su selección por identificación de la posición dactilar podría contribuir a la creación de material didáctico adaptado a personas con discapacidad visual que aporte al mejoramiento de su educación.

En la Figura 2 se ilustra un ejemplo de una imagen preprocesada, en la que hay una zona de interés identificada y la posición dactilar está ubicada en esta zona.

Figura 2

Una imagen preprocesada con una zona de interés



Estado del arte

Se pueden mencionar diversos proyectos anteriores que han utilizado modelos de inteligencia artificial para mejorar la calidad de vida de las personas con discapacidad visual. Uno de ellos es el trabajo de Montaña (2018), quien desarrolló un sistema de conocimiento del entorno que utiliza millones de imágenes para identificar el entorno de ubicación de las personas con discapacidad visual. Este sistema emplea el software OpenCV acrónimo de Open Computer Vision en una Raspberry Pi para reconocer objetos y asociarlos con un mensaje de voz, lo que ayuda a orientar a las personas en cuanto a los posibles obstáculos con los que se puedan encontrar. Aunque este sistema es útil para la navegación, no proporciona una descripción completa de la composición de la imagen, lo que limita su utilidad para la enseñanza.

Otro proyecto destacable es el trabajo de Cadena (2020), quien diseñó e implementó un prototipo de reconocimiento de objetos y asistente de navegación para personas con discapacidad visual mediante visión artificial. Este prototipo tiene modos de funcionamiento,

uno para navegar en ambientes cerrados y otro para reconocer objetos. En ambos casos, el usuario es alertado mediante mensajes de voz, lo que le permite movilizarse con autonomía propia y desarrollar mapas mentales del entorno.

Por su parte, el trabajo de Muñoz (2021) se enfoca en el reconocimiento de los elementos de un circuito eléctrico básico mediante reconocimiento de imágenes. Este proyecto emplea Redes Neuronales Convolucionales (CNN) como clasificador de entrenamiento y Redes Neuronales Artificiales para leer símbolos eléctricos de una imagen del circuito eléctrico. La aplicación genera dos archivos de texto que contienen la narración explicativa de las conexiones del circuito y la matriz de conexiones.

Finalmente, el trabajo de Mena (2022) también se enfoca en la accesibilidad de imágenes de circuitos eléctricos para personas con discapacidad visual mediante reconocimiento de la posición dactilar con ayuda de la Inteligencia Artificial. Este prototipo utiliza redes neuronales convolucionales y librerías para Python 3 (OpenCV y MediaPipe) y se ejecuta en una tarjeta Nvidia JETSON NANO. El dispositivo cuenta con tres módulos que permiten el procesamiento de objetos, la detección dactilar usando OpenCV y MediaPipe, y la síntesis de audio. Mediante el ingreso de imágenes en tiempo real vía cámara web, se realiza el reconocimiento de la posición de la mano que permite el reconocimiento de componentes eléctricos. Al igual que el trabajo de Muñoz, este proyecto se enfoca en un tema específico, en este caso, los circuitos eléctricos.

Alcance

El alcance de este proyecto consiste en desarrollar un prototipo que permita la detección de la posición del dedo índice sobre una imagen digital utilizando tecnología de inteligencia artificial y el lenguaje de programación Python en una microcomputadora. Para lograr este objetivo, se dividirá la ejecución del prototipo en tres etapas:

La primera etapa se centra en el diseño de una aplicación para crear una imagen digital procesada la cual etiqueta las partes relevantes de la misma, lo que ayudará a describir la imagen con mayor precisión.

La segunda etapa consiste en el diseño del prototipo que permita la detección de la posición del dedo índice sobre la imagen. Para ello, se adquirirán datos experimentales de imágenes previamente procesadas y se entrenarán con tecnología de inteligencia artificial y el lenguaje de programación Python en una microcomputadora. Se utilizó la biblioteca OpenCV (Domínguez, 2021) y el framework MediaPipe (Google, 2021) para la imagen y la posición del dedo, respectivamente. Además, se empleará el módulo gtts para transformar texto en voz y describir la composición de la imagen digital señalada. La construcción del prototipo incluirá la estructura de soporte, una cámara de video, un parlante y la microcomputadora.

La tercera etapa se enfocará en la captura de la ubicación dactilar sobre la imagen digital mediante la cámara de video, el procesamiento de la posición y la reproducción de un audio que describa la zona de interés señalada de la composición de la imagen.

Objetivos

Objetivos generales

Desarrollar un prototipo que brinde accesibilidad a la composición una imagen digital preprocesada mediante el seguimiento de la posición dactilar en tiempo real para personas con discapacidad visual

Objetivos específicos

- Investigar los algoritmos de IA a emplearse en los procesos de detección de la posición dactilar.

- Definir módulos, bibliotecas y librerías disponibles para microcomputadores asequibles que permitan la detección dactilar, procesamiento de imágenes y reproducción texto a voz.
- Implementar una aplicación que permita preprocesar una imagen digital para etiquetar sus regiones de interés.
- Desarrollar la aplicación que realice la detección dactilar sobre la imagen y reproduzca el audio descriptivo.
- Evaluar el desempeño del prototipo con la participación de personas con discapacidad visual.

Organización del trabajo

El trabajo de titulación está constituido en cinco diferentes capítulos que describen el desarrollo del prototipo que brinde accesibilidad a la composición de una imagen digital preprocesada mediante el seguimiento de la posición dactilar para personas con discapacidad visual. El capítulo I se describe la introducción, justificación e importancia, estado del arte, objetivos generales y específicos. El capítulo II se encuentra el marco teórico donde el lector podrá conocer y entender el trabajo de titulación, tales temas como las oportunidades tecnológicas de personas con discapacidad visual, el procesamiento de imágenes digitales, la descripción y clasificación del mundo de la Inteligencia Artificial para lograr utilizar sus campos y subcampos para el desarrollo del prototipo y también la especificación del software y hardware para la creación del dispositivo que beneficiará para la accesibilidad a imágenes según la posición dactilar. El capítulo III comprende de la ejecución del prototipo donde se evidenciará los materiales y metodología describiendo la tecnología, librerías y algoritmos utilizadas según la necesidad del caso, como también la configuración y desarrollo del hardware y software. El capítulo IV y V consta del análisis de las pruebas y resultados de las mismas con sus respectivas conclusiones, recomendaciones y trabajos futuros.

Capítulo II

Marco teórico

Discapacidad visual

La discapacidad visual es un desafío de vida para las personas que lo presentan, experimentando la pérdida parcial o total de la capacidad para ver, afectando realizar actividades tanto cotidianas, laborales y educativas. Sin embargo, estas personas desarrollan habilidades significativas adaptándose físicamente a la discapacidad visual permitido por el proceso de neuroplasticidad que significa la capacidad que tiene el cerebro para reestructurar y naturalizar las nuevas situaciones o cambios en su entorno (Lewis, 2019).

Las oportunidades se limitan por la falta de desarrollo de las habilidades para su adaptación, tanto que, al incluir recursos y tecnología estas aportan exponencialmente a que lleven una vida independiente y productiva. Las personas no videntes al interactuar con gráficos permiten la transmisión de información y amplia el conocimiento, que si bien es cierto era complejo años atrás, pero en la actualidad hay tecnología inclusiva para la accesibilidad a los gráficos. Esto puede incluir dispositivos como: lectores de pantalla, gráficos táctiles, materiales en braille e interfaces hápticas para el acceso a educarse.

Las investigaciones han demostrado que el uso de tecnologías de asistencia ha mejorado la accesibilidad de los gráficos para las personas con discapacidad visual. Por ejemplo, en el estudio realizado por McGookin et al. (2010) evaluaron la efectividad de una interfaz háptica para transmitir información sobre elementos gráficos. El estudio encontró que los participantes pudieron identificar elementos gráficos de manera más precisa y eficiente con la interfaz háptica que con una maqueta física. Otro estudio de (Kuriakose et al., 2022) investigaron el uso de retroalimentación audio-táctil para hacer que los gráficos interactivos fueran accesibles para personas con discapacidad visual. El estudio encontró que el uso de

comentarios táctiles de audio mejoró significativamente la capacidad del usuario para explorar y comprender la información gráfica.

Este trabajo pretende solucionar los problemas existentes que dificultan a las personas no videntes a la interacción con gráficos con el fin del desarrollo de una herramienta eficiente para la accesibilidad a la composición de una imagen mediante el seguimiento de la posición dactilar que finalmente transmita la información en un audio.

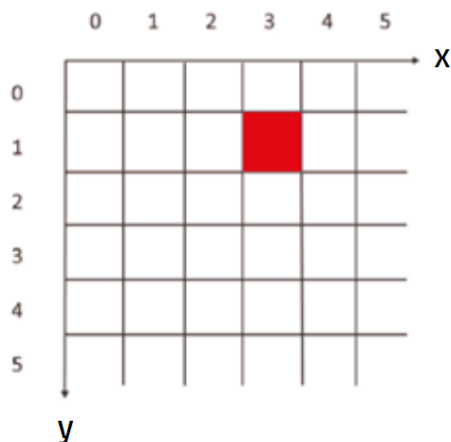
Procesamiento digital de imágenes

Las imágenes son representaciones visuales de objetos o escenas capturadas mediante dispositivos electrónicos y son percibidos por la vista. Estas imágenes en el contexto de procesamiento digital de imágenes se componen de una serie de elementos básicos, como los *píxeles*, abreviatura de su denominación inglesa *picture elements*, que se organizan en una matriz bidimensional de intensidad de luz $f(x, y)$, donde x e y representan las coordenadas espaciales y la unión de varios puntos forman la imagen completa (Gonzalez & Woods, 2002).

Las coordenadas espaciales de la imagen se definen en términos de una cuadrícula. La coordenada x se extiende horizontalmente de izquierda a derecha, mientras que la coordenada y se extiende verticalmente de arriba hacia abajo. Cada píxel en la imagen tiene una coordenada única (x, y) cómo se ilustra en la Figura 3 (Domínguez, 2021).

Figura 3

Coordenadas espaciales



Nota. Tomado de *Visión artificial: aplicaciones prácticas con OpenCV-Python (Unidad 3)*, por Mínguez, T. D., 2021, Marcombo.

Una imagen digitalizada puede encontrarse en escala de grises o en color. En el caso de una imagen en escala de grises, cada píxel tiene un valor numérico que representa la intensidad de gris en ese punto. En el caso de una imagen en color, cada píxel tiene tres valores numéricos que representan la intensidad de los componentes rojo, verde y azul (RGB) en ese punto.

El procesamiento digital de imágenes incluye técnicas para analizar y manipular información visual extrayendo características relevantes que son aplicativas para la orientación, redimensionamiento y transformaciones geométricas de imágenes.

Orientación

La inversión de la imagen verticalmente y horizontalmente permite la visualización en modo espejo permitiendo a aplicaciones comparaciones simétricas, correcciones de dimensiones y calibración de la posición de la cámara.

Redimensión

Una imagen permite el redimensionamiento de su tamaño al deseado, lo que implica el ajuste de su alto y ancho, esta técnica se utiliza con el fin de optimizar el tiempo y preparación para el procesamiento para que se realice de manera uniforme estandarizando dimensiones para un ajuste de tamaño visual que se vaya a ser mostrado al usuario.

Transformación de perspectiva

Esta técnica permite la transformación geométrica de un área específica en una vista de perspectiva definida por puntos de referencia tanto que se muestre como una imagen tomada perpendicularmente al plano conservando sus propiedades como se visualiza en la Figura 4, esta transformación se lo realiza a partir de una matriz de transformación que permite rectificar la imagen para lograr tener en un plano, es decir aplica la teoría de transformación lineal, lo que implica mapear las coordenadas de la imagen no colineales original a una nueva ubicación en la imagen de cambio de perspectiva manteniendo el paralelismo de las líneas y contornos de la imagen (López de Paz, 2012). Los puntos de referencia a elegirse es de al menos cuatro vértices de modo que la posición x y y son identificados en píxeles. (Aguirre & Olmedo, 2016)

La matriz que define López de Paz (2012) a la transformación de perspectiva es de orden (3x3) denominada matriz homografía H_s .

$$\begin{bmatrix} x_i \\ y_i \\ w \end{bmatrix} = H_s \begin{bmatrix} x_w \\ y_w \\ W \end{bmatrix}$$

El sistema de coordenadas del plano original es $(x_w, y_w, W)^T$; $z_w = 0$ mientras que el sistema de coordenadas del nuevo plano es $(x_i, y_i, w)^T$.

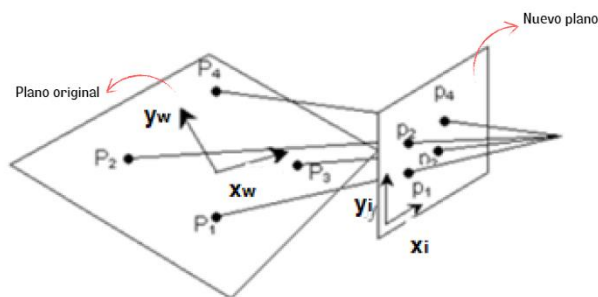
x_w, y_w y W representan las coordenadas en el plano original (plano 3D), mientras que x_i

e y_i representan las coordenadas en el nuevo plano de la imagen (plano 2D), W se refiere a la coordenada z en el espacio tridimensional. Para simplificar los cálculos, generalmente se

establece $z = 0$ en el sistema de referencia original, lo que implica que $W = 0$. Esto se debe a que las coordenadas en el plano de la imagen (plano 2D) se proyectan en un plano paralelo al plano xy del plano original.

Figura 4

Transformación de perspectiva



Nota. Planos en una transformación de perspectiva. Adaptado de *Diseño de un programa de ortorectificación y georreferenciación de imágenes aéreas aplicadas a campos de caña de azúcar* (p.17), por López de Paz, G. J., 2012, Pontificia Universidad Católica del Perú.

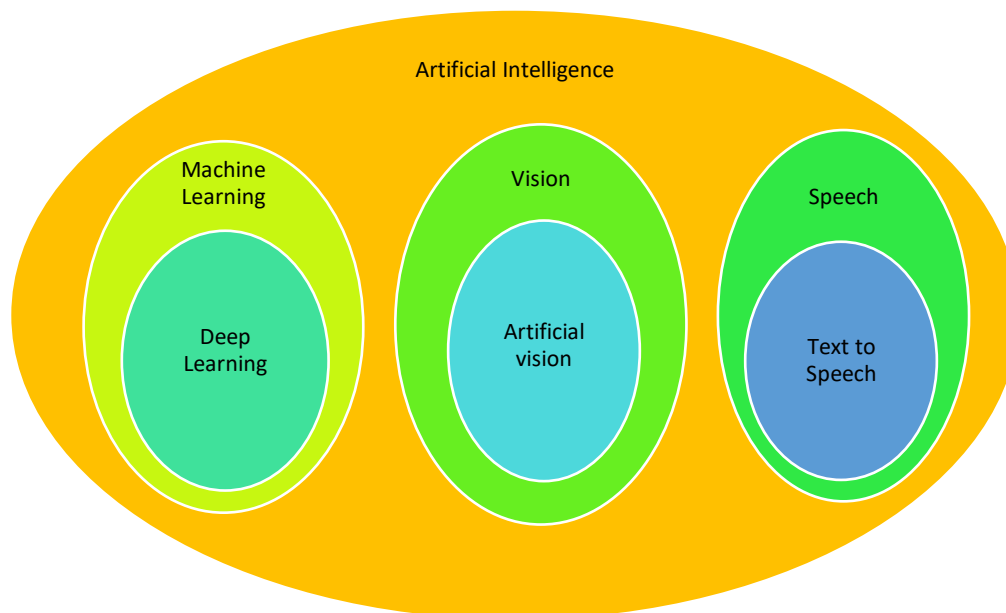
Inteligencia Artificial

Es una tecnología que se ha introducido en el mundo y lo utilizan a diario para el desarrollo de grandes dispositivos que permitan imitar fácilmente la inteligencia humana y la realizar actividades específicas. Bellman, un matemático definió a la inteligencia artificial cómo la automatización de actividades asociadas a las habilidades de pensamiento humano incluyendo la toma decisiones y aprendizaje (Khanagar et al., 2021).

El desarrollar dispositivos potenciales que puedan aprender a través de datos para resolver problemas es el objetivo de todo lo que abarca la inteligencia artificial como se ilustra en la Figura 5.

Figura 5

Campos dentro de la Inteligencia Artificial



Machine Learning

El aprendizaje automático es el subcampo de la Inteligencia Artificial, que se centra en algoritmos basados en un conjunto de datos para aprender y predecir resultados. El propósito del aprendizaje automático es que un dispositivo pueda resolver problemas sin intervención. Este subcampo ha revolucionado sobre la visión artificial y la robótica.

El aprendizaje automático se clasifica en:

- Aprendizaje supervisado
- Aprendizaje no supervisado
- Aprendizaje por refuerzo.

En el aprendizaje supervisado, el algoritmo aprende con la experiencia, tal caso como proporción de un conjunto de datos reales conocidos para el entrenamiento, datos de muestra de entrada y salida que según el criterio este algoritmo pueda predecir y se pone a prueba la

capacidad de la predicción. (Anzola, 2022) En el aprendizaje no supervisado, el algoritmo recibe datos sin información previa y aprende a encontrar patrones o tendencias en los datos. (Álvarez Vega et al., 2020). Y por último el aprendizaje por refuerzo, el algoritmo aprende a prueba y error, tanto de datos etiquetados y sin etiquetar e interactúan entre sí, logrando la retroalimentación en forma de perfeccionamiento para obtener mejores predicciones, caracterizaciones o clasificaciones. (Álvarez Vega et al., 2020)

Deep Learning. El aprendizaje profundo es un componente del aprendizaje automático que utiliza la red con diferentes capas computacionales en una red neuronal profunda para analizar los datos de entrada modelando la imitación de cómo funciona el cerebro humano para remodelar el futuro de la Inteligencia Artificial. A diferencia de las redes neuronales el aprendizaje profundo hace el uso de muchas neuronas y capas ocultas. Estas capas ocultas son capaces de extraer automáticamente características relevantes de los datos de entrada y utilizarlas para la clasificación, detección de objetos, segmentación de imágenes, entre otras tareas. A diferencia del aprendizaje automático tradicional, que se basa en características manuales seleccionadas por los ingenieros de datos, el aprendizaje profundo aprende automáticamente las características y las relaciones entre los datos de entrada, lo que lo hace muy efectivo en una amplia variedad de aplicaciones (Ravi et al., 2017).

Las redes neuronales es la base del aprendizaje profundo y son un conjunto de algoritmos del aprendizaje automático inspirados en las redes neuronales biológicas con el propósito que funcionen como el cerebro humano basados en principios computacionales. Las redes neuronales están compuestas de un conjunto de neuronas artificiales que se organizan por capas y se interconectan mediante conexiones sinápticas. Estas conexiones, que se denominan pesos, son ajustadas por un proceso de aprendizaje que permite que la red pueda realizar tareas específicas. (Zador, 2019)

Existen varios tipos de redes neuronales según la arquitectura del aprendizaje profundo, pero las más comunes son:

- Redes neuronales feedforward
- Redes neuronales recurrentes
- Redes neuronales convolucionales

Las redes neuronales feedforward son aquellas en las que la información fluye en una sola dirección, desde la entrada hasta la salida, sin retroalimentación esto hace que nunca la información pueda ser transmitida más de una vez a través de una neurona antes de dar una salida como respuesta. (Vásquez, 2014) Las redes neuronales recurrentes, por otro lado, tienen conexiones de retroalimentación, lo que les permite procesar secuencias de datos. Las redes neuronales convolucionales son especializadas en procesar datos de imágenes y se basan en la operación de convolución que es una forma fácil de realizar operaciones complejas utilizando el filtro de convolución. Estas no utilizan núcleos predefinidos, sino que aprende neuronas conectadas localmente que representan núcleos específicos de datos. Dado que estos filtros se aplican repetidamente a toda la imagen (Ravi et al., 2017).

Si bien es cierto las redes neuronales convolucionales se destaca por los avances significativos en temas como el procesamiento de imágenes tales como la detección de objetos, reconocimientos de gestos o emociones y el seguimiento de partes del cuerpo.

Artificial Vision

La visión artificial es un campo de la inteligencia artificial con el objetivo de crear sistemas que puedan extraer y descifrar información visual de imágenes y videos, permitiendo que los dispositivos tomen decisiones basadas en lo que captan tales como la clasificación y búsqueda. La visión artificial puede aplicarse en diferentes ámbitos cómo la medicina, industria, comercio, seguridad y entre otros que está ganando terreno (Domínguez, 2021).

Existen librerías que ocultan su complejidad matemática de los algoritmos para realizar dicho trabajo cómo lo es OpenCV, lo cual las aplicaciones se observarían en tiempo real.

Detección de la posición dactilar

La detección de la posición dactilar se refiere a la identificación y localización de algún rasgo de la estructura de la mano, este proceso implica la segmentación y el preprocesamiento de la imagen, la detección de características clave y la extracción de puntos de interés. Existen varios algoritmos de inteligencia artificial (IA) que son utilizados para este propósito, tales como:

- **Redes neuronales convolucionales (CNN):** Es un algoritmo eficaz para detectar y clasificar de una imagen los dedos de una mano. Estas redes neuronales pueden aprender automáticamente características relevantes y realizar la identificación y localización precisa de los dedos para la detección de la posición dactilar (Intel Corporation, s. f.).
- **Modelos de detección de objetos:** Algunos modelos como YOLO (You Only Look Once) o SSD (Single Shot MultiBox Detector), son ampliamente utilizados para la detección y localización de objetos en imágenes. Estos modelos se pueden entrenar para reconocer y seguir los dedos de una mano, proporcionando información sobre su posición dactilar en la imagen (Dewi & Juli Christanto, 2022).
- **Redes neuronales recurrentes (RNN):** Es un algoritmo capaz de analizar las características secuenciales de una secuencia de imágenes en tiempo real y estimar la posición dactilar actual en cada imagen.
- **Métodos basados en marcadores o puntos de referencia:** Se basa en el uso de marcadores o puntos de referencia colocados en los dedos de una mano. Los algoritmos de IA pueden utilizar técnicas de detección y seguimiento de puntos

clave para localizar y estimar la posición dactilar en función de los marcadores identificados (Jara et al., 2019).

Python

Es un lenguaje de programación más utilizado en el mundo con un mayor número de programadores por la sencillez de aprendizaje y facilidad de uso que aporta. Este lenguaje incluye programación imperativa cómo también programación orientada a objetos, además posee un entorno multiplataforma para diferentes sistemas operativos y entre ellos la Raspberry (Domínguez, 2021).

En base a Python (2023) existen algunos tipos de librerías útiles tanto para Deep Learning, Machine Learning, cálculos numéricos, visualización, inteligencia artificial y procesamiento de lenguaje natural tales como las siguientes bibliotecas:

- **OpenCV:** OpenCV es una biblioteca de procesamiento de imágenes y visión artificial de código abierto. Proporciona herramientas y algoritmos para procesar y analizar imágenes y videos en tiempo real, actualmente es una de las herramientas más utilizadas en el desarrollo de aplicaciones de visión. Las interfaces de calculadora y programación están disponibles en diferentes lenguajes como C ++, Java y Python, así como en las plataformas Windows, Linux, Mac OS y Android (Domínguez, 2021). OpenCV ayuda con varias tareas, como el procesamiento de imágenes y videos, la detección de objetos, el aprendizaje automático y la visión 3D. OpenCV cubre el procesamiento de imágenes y videos en el sentido de proporcionar herramientas de procesamiento de imágenes y videos, como filtrado, transformación de geometría, detección y seguimiento de los bordes de objetos y características específicas, como líneas y círculos en diferentes formatos (Howse, 2013). Además, OpenCV contiene herramientas para crear y entrenar modelos de aprendizaje automático, incluidos algoritmos de clasificación, regresión y segmentación de

imágenes. Otra herramienta de OpenCV que se enfoca en la reconstrucción de escenas 3D a partir de imágenes estéreo y calibración de cámaras (Quintana et al., 2012).

- **MediaPipe:** MediaPipe es una biblioteca desarrollada por Google que proporciona un conjunto de plantillas y herramientas previamente entrenadas para el procesamiento de medios en tiempo real, permite la construcción de aplicaciones de visión por computadora con el uso de técnicas de Deep Learning y que va expandiéndose al transcurrir el tiempo, incluyendo redes neuronales convolucionales, para lograr un rendimiento de alta precisión, velocidad y extraer características relevantes de las imágenes y los videos con él, puede realizar tareas como reconocimiento de manos, reconocimiento de rostros y estimación de ubicación (Zhang et al., 2020).
- **ArUco:** ArUco es una biblioteca de detección de marcadores visuales utilizada en aplicaciones de realidad aumentada y seguimiento de objetos. Permite detectar y reconocer marcadores predefinidos en imágenes y vídeos, lo que facilita la estimación precisa de la posición y orientación de la cámara. ArUco es ampliamente utilizado para crear experiencias interactivas y precisas en entornos virtuales y aumentados (Jara et al., 2019).
- **NumPy:** NumPy es una biblioteca básica de Python para manipular y calcular matrices numéricas. Ofrece una gran cantidad de funciones y operaciones para trabajar de manera eficiente con datos digitales.
- **Pyzbar:** Pyzbar es una biblioteca de Python para reconocer y decodificar códigos de barras y códigos QR en imágenes y videos. Facilita la extracción de información codificada en estos formatos.
- **gTTS:** gTTS (Google Text-to-Speech) es una biblioteca de síntesis de voz. Proporciona una interfaz simple para generar archivos de audio a partir de texto. El sintetizador de texto a voz (TTS) puede leer cualquier texto escrito y transformarlo a voz, tal que la

voz será reproducida mediante la traducción automática de las palabras a pronunciar. Este método es muy usado para ayuda de personas con discapacidad visual cómo también en películas de ciencia ficción y automatización de centros de llamadas entre otros, en Taylor (2009) menciona que Stephen Hawking probablemente haya contribuido más que nadie a un tipo particular de voz sintética. La voz sintética es una limitante para la tecnología TTS porque puede sonar robotizada y poco natural que en algunos casos dificulte la comprensión, pero de una u otra manera su beneficio es amplio tanto para la inclusividad, comunicación y entretenimiento.

- **Playsound:** Playsound es una biblioteca de Python que proporciona una manera fácil de reproducir archivos de audio. Con él puedes reproducir sonidos y música de forma rápida y directa.
- **Shapely:** Shapely es una biblioteca para analizar y manipular geometría plana. Proporciona herramientas para realizar operaciones espaciales y calcular relaciones entre objetos geométricos.
- **Tkinter:** Tkinter es la biblioteca estándar de Python para crear interfaces gráficas de usuario (GUI). Proporciona componentes y herramientas para crear aplicaciones con ventanas, botones, cuadros de texto y otros elementos interactivos.
- **PIL:** PIL (Biblioteca de imágenes de Python) es una biblioteca de imágenes de Python. Permite operaciones como abrir, editar y guardar imágenes en diferentes formatos.
- **Pandas:** Pandas es una biblioteca de Python para el análisis de datos. Proporciona estructuras de datos flexibles y potentes como DataFrame para administrar y manipular información tabular.
- **qrcode:** qrcode es una biblioteca de Python para generar códigos QR. Con él, puede crear fácilmente códigos QR a partir de texto, enlaces, contactos y otros tipos de información.

- CVS: CVS es un popular sistema de control de versiones para rastrear cambios en archivos y proyectos de software.
- Reportlab: Reportlab es una biblioteca de Python para generar documentos PDF. Proporciona herramientas y funciones para crear informes, gráficos, etc. en formato PDF.
- Imutils: Imutils es una biblioteca de Python que proporciona una colección de funciones y herramientas para ayudarlo a procesar imágenes. Le ayuda a realizar de manera eficiente tareas comunes como cambiar el tamaño, rotar y editar imágenes.

Raspberry Pi

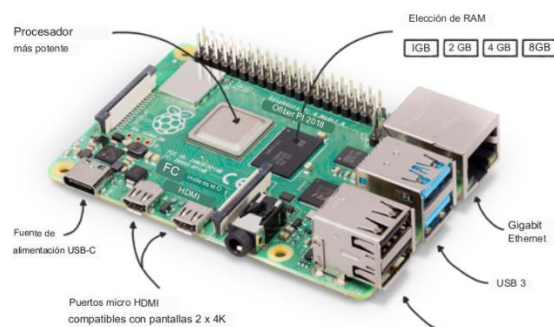
Es una microcomputadora de bajo coste con la accesibilidad a la integración con dispositivos externos desarrollada por la Fundación Raspberry Pi. Esta microcomputadora integra un CPU con arquitectura ARM, GPU, memoria RAM de 2,4,8 GB y una serie de periféricos de entrada y salida. La Raspberry es una plataforma versátil ya que admite muchos lenguajes de programación diferentes a Python, como por ejemplo C, Ruby, Java y Perl e incluso posee un lenguaje de programación propio con su entorno de desarrollo denominado Pi para crear música (Richardson & Wallace, 2016) como también se puede ejecutar una variedad de sistemas operativos, el software es compatible con anteriores versiones de Raspberry Pi en cualquier modelo anterior.'

Raspberry Pi 4 Model B

Existen varias versiones de Raspberry Pi tanto cómo la 1,2,3 y la actual 4, que salió al mercado en el 2019. La Raspberry Pi 4 Model B es el modelo principal de su versión porque posee características más completas tales como se ilustra en la Figura 6.

Figura 6

Microcomputadora Raspberry Pi 4 Model B



Nota. La Figura representa la integración de periféricos en la microcomputadora Raspberry 4 Model B (Raspberry Pi, 2019)

Tabla 1

Especificaciones técnicas de la Raspberry Pi 4 Model B

Raspberry Pi 4 Modelo B	Características
8GB	
Procesador	Broadcom BCM2711, SoC de cuatro núcleos Cortex-A72 (ARM v8) de 64 bits a 1,5 GHz
Memoria	8GB
Conectividad	2,4 GHz y 5,0 GHz IEEE 802.11b/g/n/ac inalámbrico LAN, Bluetooth 5.0, BLE Gigabit Ethernet 2 x puertos USB 3.0, 2 x puertos USB 2.0
Compatibilidad con tarjeta SD	Ranura para tarjeta Micro SD para cargar el sistema operativo y el almacenamiento de datos
Ambiente	Temperatura de funcionamiento 0–50°C

Raspberry Pi 4 Modelo B	Características
8GB	
<hr/>	
GPIO	Encabezado GPIO estándar de 40 pines
Video y sonido	2 × puertos micro HDMI (compatible con hasta 4Kp60) Puerto de pantalla MIPI DSI de 2 carriles y Puerto de cámara MIPI CSI de 2 carriles Puerto de audio estéreo y video compuesto de 4 polos
Multimedia	H.265 (descodificación 4Kp60) H.264 (decodificación 1080p60, codificación 1080p30) OpenGL ES, gráficos 3.0

Cámara

Dispositivo que permite la captación de imágenes full HD en tiempo real de alta definición y calidad de la misma.

Figura 7

Cámara web



Tabla 2*Especificaciones técnicas de la cámara*

Cámara web	Características
Resolución y calidad	720p, 30fps
Resolución dinámica	640*480
Tipo de enfoque	Foco fijo
Campo visual	60°
Conexión	Alámbrica con cable puerto USB

Capítulo III

Metodología de desarrollo del proyecto

La limitación a la accesibilidad de imágenes para las personas con discapacidad visual es contundente, actualmente existe dificultades para cumplir tareas que requieren de interacción gráfica. No es lo mismo solo escuchar la descripción de la composición de una imagen, si a ésta adicionalmente se sincroniza según la posición dactilar que interactúe con la zona de interés, con el fin de que pueda sentir, escuchar e imaginar, es decir, que tenga una percepción háptica y cognitiva de la información que la imagen le proporciona.

Por ende, se busca brindar a las personas con discapacidad visual la oportunidad de tener una forma fácil e intuitiva de interactuar con imágenes y también conocer su composición auxiliados con el seguimiento de la posición dactilar, utilizando todos los sentidos hábiles para que pueda tener una experiencia única y personalizada.

La necesidad de este proyecto recae en que el no vidente interactúe con una imagen tanto que para concretar el desarrollo del prototipo se requirió el medio computacional tal como el microcomputador Raspberry Pi 4 y hardware adicional que consta de la estructura de soporte, una cámara de video y un altavoz. La ejecución del aplicativo se enfoca en la detección dactilar sobre la imagen tal que para ello se realizó la adquisición de datos de manera experimental en imágenes preprocesadas (nombre de la imagen, región de interés de la imagen y coordenadas de la imagen) para posteriormente entrenarlas en el lenguaje de programación Python en una microcomputadora y asociarlas en tiempo real con la posición dactilar detectada, donde se utiliza la biblioteca OpenCV y el framework MediaPipe que ayuda tanto para la imagen y el seguimiento de la posición dactilar respectivamente, adicional el módulo gtts que transforma el texto a voz para describir la composición de la imagen digital señalada. En fin, se realiza la captación con la cámara de video que detectará la ubicación dactilar sobre la imagen preprocesada, esta posición es comparada con la base de datos

creada y según la información encontrada se reproducirá un audio que describirá la zona de interés de la imagen.

Configuración de la Raspberry Pi 4 Model B

La configuración e instalación de los recursos para que el hardware del prototipo funcione de manera óptima en la microcomputadora, posterior a un análisis de compatibilidad entre sistemas operativos y bibliotecas, se establece la descarga del sistema operativo de escritorio hereditario de la Raspberry Pi denominado Raspbian Debian Linux 10 Buster, donde la imagen se puede adquirir de: <https://www.raspberrypi.com/software/operating-systems/#raspberry-pi-os-legacy> que posterior lo instalará en la tarjeta SD para arrancar la Raspberry Pi y se puede verificar la correcta instalación cómo se ilustra en la Figura 8.

Figura 8

Información del Sistema Operativo de la Raspberry Pi 4

```
pi@raspberrypi:~$ cat /etc/os-release
PRETTY_NAME="Raspbian GNU/Linux 10 (buster)"
NAME="Raspbian GNU/Linux"
VERSION_ID="10"
VERSION="10 (buster)"
VERSION_CODENAME=buster
ID=raspbian
ID_LIKE=debian
HOME_URL="http://www.raspbian.org/"
SUPPORT_URL="http://www.raspbian.org/RaspbianForums"
BUG_REPORT_URL="http://www.raspbian.org/RaspbianBugs"
```

El sistema operativo Raspbian Debian Linux 10 Buster incluye por default el lenguaje de programación Python 2 y Python 3, Python 2 está fuera de su rango de vida útil por déficit de actualizaciones de seguridad tanto que se están migrando proyectos a la versión predeterminada que es Python 3 dando a notar un lenguaje más entendible y legible con gran visión incluyendo y actualizando diversas herramientas a las bibliotecas, recomendada para uso a proyectos de desarrollo de Inteligencia Artificial como es en este caso esencial para el desarrollo del software.

Bibliotecas

Existe una amplia gama de bibliotecas en Python 3 que permite el desarrollo de diversas aplicaciones, estas bibliotecas ya poseen módulos, paquetes, *frameworks* que dentro de ellas existe un conjunto de funciones ya incorporadas para facilitar el esfuerzo al programar y el ahorro de tiempo, proveen soluciones estandarizadas (Python, 2023).

Las bibliotecas se instalan con la siguiente instrucción en el terminal de comandos `pip3 install <biblioteca>`, si bien el uso de las bibliotecas proporcionó herramientas para el preprocesamiento de imágenes, generación de códigos QR cómo también decodificación, generación de voz sintética, manipulación de datos, interfaces gráficas, generación de PDF y herramientas para el uso de objetos geométricos y análisis de las mismas.

Diagrama general del aplicativo

La aplicación desarrollada para preprocesar una imagen digital y etiquetar las regiones de interés es esencial en este apartado porque es ahí donde adquirimos la base de datos para ejecutar el reconocimiento de la zona de interés según la posición dactilar sobre una imagen interactuando en tiempo real. Este aplicativo combinó diferentes funcionalidades tales como el procesamiento de imágenes, detección de una mano y reconocimiento de marcadores visuales.

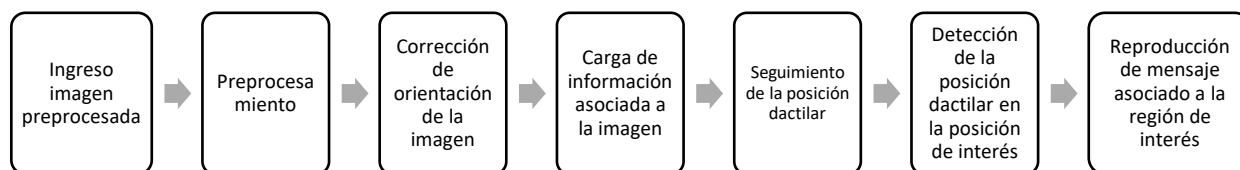
Las etapas del desarrollo del aplicativo permiten la accesibilidad a la composición de una imagen digital preprocesada mediante el seguimiento de la posición dactilar para personas con discapacidad visual son las que se ilustran en el diagrama de bloques de la **¡Error! No se encuentra el origen de la referencia..**

Las personas con discapacidad visual son capaces de experimentar el aplicativo que permite reconocer la composición de una imagen con una percepción háptica, auditiva y cognitiva para la comprensión de la información, el sistema está desarrollado con el fin de fortalecer las capacidades de autonomía al enfrentarse al reconocimiento de imágenes y la

composición, simplemente esta habilidad se desarrolla en el cerebro y es capaz de percibir información sensorial no visual de manera eficiente, este aplicativo de asistencia permite la construcción de la representación mental de la imagen percibida por los sentidos funcionales.

Figura 9

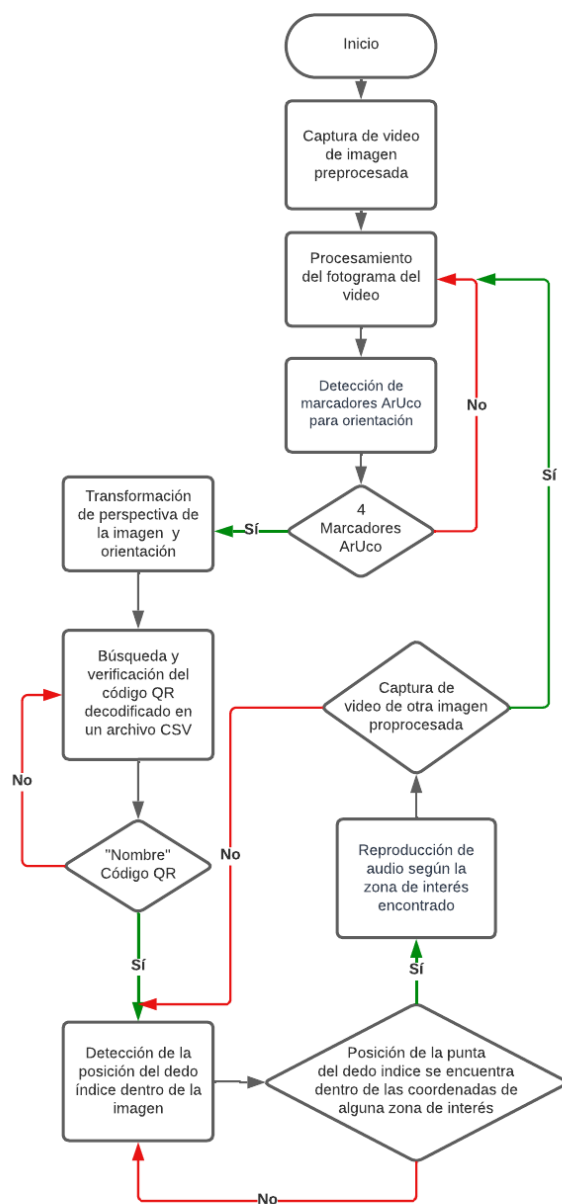
Diagrama de bloques de la aplicación de la descripción auditiva de la zona de interés según la posición



El propósito del aplicativo es en asistir a la persona con discapacidad visual, tanto que, el prototipo se construyó con un microcomputador, cámara y reproductor de sonido. En el microprocesador está contenido de su programación, la cámara capta la imagen a ser percibida cómo también el dedo índice que localiza un punto, que de estar contenido en la posición de una zona de interés el reproductor de sonido lo describe, esto se procesa en tiempo real y se realiza como se ilustra en la Figura 10.

Figura 10

Diagrama de flujo del aplicativo de la descripción auditiva de la zona de interés según la posición dactilar sobre una imagen

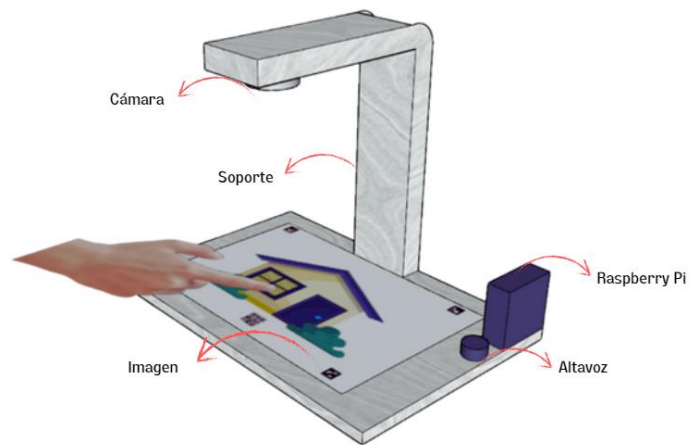


Ingreso de imagen preprocesada

La captura de video se lo realiza sobre una imagen preprocesada con una cámara como se ilustra en la Figura 11, Para garantizar una captura precisa y cómoda de la imagen y la mano, se ajusta la altura de la cámara a 45 cm sobre la superficie. Este ajuste estratégico define la posición óptima. El proceso de preprocesamiento utiliza la biblioteca de OpenCV, la que continuamente permite el procesamiento de cada imagen individual en la secuencia del video, alguna de las operaciones que se realiza dentro del procesamiento de la imagen es la obtención de la altura, el ancho y el giro en modo espejo de la imagen. La imagen procesada se muestra en una ventana en forma de video en tiempo real.

Figura 11

Ingreso de imagen preprocesada



4

(a) Modelamiento del ingreso de imagen preprocesada



(b) Ingreso de imagen preprocesada en tiempo real

Preprocesamiento

El propósito del preprocesamiento es obtener una base de datos de imágenes procesadas que contengan información sobre las coordenadas y la región de interés asociada a un mensaje de voz. Esta etapa base del *prototipo* se encarga de establecer la composición de la imagen, permitiendo ajustar su tamaño, resolución y formato para estandarizar y normalizar todas las imágenes a etiquetarse. En el proceso, se resaltan las regiones de interés.

En el diagrama de bloques del preprocesamiento, ilustrado en la Figura 12, se muestra cómo se lleva a cabo esta etapa. Para etiquetar las regiones de interés, se propone que un asistente de la persona con discapacidad visual seleccione previamente las imágenes según la necesidad. Luego, sobre cada imagen seleccionada, se dibujan las regiones de interés contorneándolas con puntos al hacer clic con el mouse. Automáticamente se forma el polígono

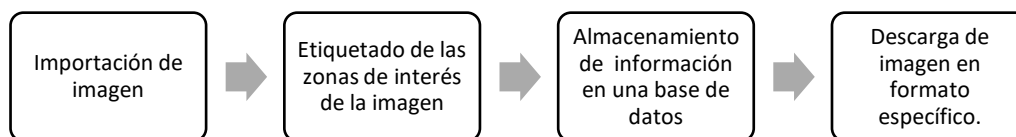
correspondiente a la región de interés. Una vez finalizado el reconocimiento de la zona, se ingresa el nombre del etiquetado y se guarda.

Internamente, se genera un archivo .csv que almacena la información obtenida, creando así una base de datos para su uso posterior. Además, esta aplicación genera un PDF que incluye marcadores de ArUco. Estos marcadores proporcionan una orientación y referencia precisa para que una cámara pueda reconocerlos en tiempo real. Además, el PDF descargado incluye la imagen preprocesada y un código QR para la identificación de la imagen.

Posteriormente, el usuario puede imprimir esta imagen, obteniendo así el material necesario para su uso.

Figura 12

Diagrama de bloques del preprocesamiento de una imagen



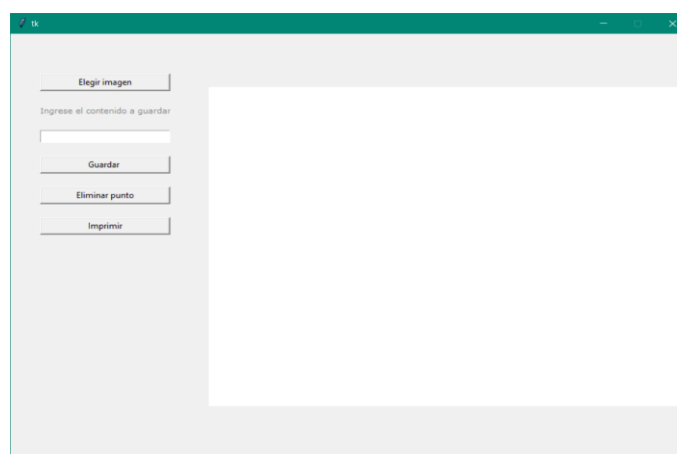
La interfaz gráfica del preprocesamiento se ha implementado utilizando la biblioteca Tkinter en Python. Se eligió Tkinter debido a su compatibilidad con diversas necesidades de visualización en pantalla. Permite controlar el diseño de la aplicación de manera geométrica, facilitando la ubicación de ventanas y *widgets*, que son elementos como botones y cuadros de texto que se encuentran contenidos dentro de una ventana. La interfaz desarrollada con Tkinter ha demostrado ser estable y funcional en la actualidad (Amos, 2020).

La interfaz gráfica está diseñada de acuerdo a lo ilustrado en la Figura 13 consta de una ventana principal que utiliza clases de Tkinter, como el *canvas*, para mostrar la imagen seleccionada, y el *frame*, que actúa como contenedor de los diferentes *widgets*.

Entre los widgets utilizados se encuentra el botón "Elegir imagen", el cual permite cargar archivos de imágenes con las extensiones *.jpg*, *.jpeg* o *.png*. Además, se ha agregado un campo de texto para ingresar la etiqueta correspondiente a la región de interés, un botón para guardar los datos, otro para eliminar el último punto seleccionado y, por último, el botón "imprimir".

Figura 13

Interfaz de la aplicación para el preprocesamiento de la imagen

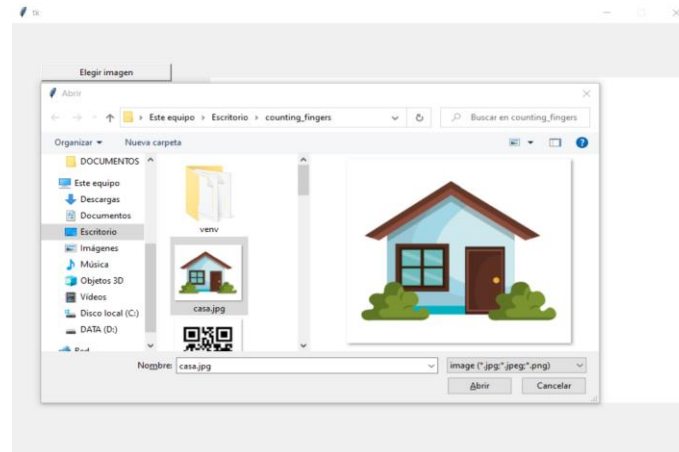


Importación de imagen. Es una función creada para que el usuario que asistirá a la persona con discapacidad visual pueda seleccionar una imagen en un cuadro de diálogo en el computador para que seleccione el archivo como se visualiza en la Figura 14. Este archivo seleccionado es leído utilizando la biblioteca OpenCV y se redimensiona a una resolución de 640×480 *pixeles*, esta se convierte en un objeto de imagen usando la biblioteca *pillow*, lo que posibilita la visualización en pantalla y en ella su interacción.

Además, durante este proceso, la imagen seleccionada se guarda en formato *.png* y se genera un código QR basado en el nombre del archivo seleccionado. De esta manera, se facilita la identificación y referencia futura de la imagen.

Figura 14

Importación de imagen a preprocesar



Etiquetado de las zonas de interés de la imagen. Tiene como objetivo limitar y definir áreas específicas en una imagen. Para lograr esto, se utilizan funciones de eventos del mouse que permiten crear puntos al hacer clic en la imagen. Estos puntos se almacenan coordenadas x e y y se actualiza una lista de puntos a medida que se van creando.

Los puntos etiquetados se visualizan como óvalos en el lienzo de la imagen elegida, como se muestra en la Figura 15. Conforme se crean nuevos puntos, se interconectan automáticamente mediante una función conexión de puntos, formando así un polígono. Es importante destacar que la función de conexión de puntos siempre cierra el polígono al unir el último punto con el primero.

De esta manera, se logra definir y delimitar las zonas de interés en la imagen, lo que permite identificar las áreas específicas que se desean destacar. Estas zonas de interés son importantes para el procesamiento posterior de la imagen y el análisis de la información relevante contenida en ellas.

Figura 15

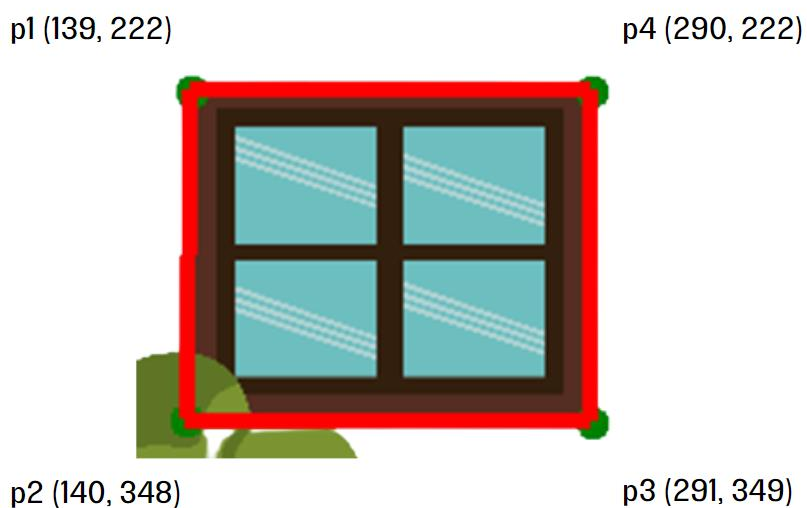
Selección de puntos en la imagen de la zona de interés



Conexión de puntos. Es una función creada se encarga de trazar líneas que conectan los puntos almacenados, utilizando parámetros como el tipo de widget en el que se dibujará y el color de las líneas. Esta función desempeña un papel fundamental en la formación del polígono que define la zona de interés. Al recibir los puntos almacenados, la función utiliza un ciclo repetitivo para trazar las líneas que conectan cada par de puntos consecutivos. Como se muestra en la Figura 16,, este proceso de conexión de puntos resulta en la formación de un polígono, el cual delimita la zona de interés deseada.

Figura 16

Formación del polígono de la zona de interés



En el ejemplo ilustrado, se ha formado un polígono al seleccionar consecutivamente cuatro puntos: p1, p2, p3 y p4. Esta secuencia de puntos define los límites de la zona de interés en la imagen.

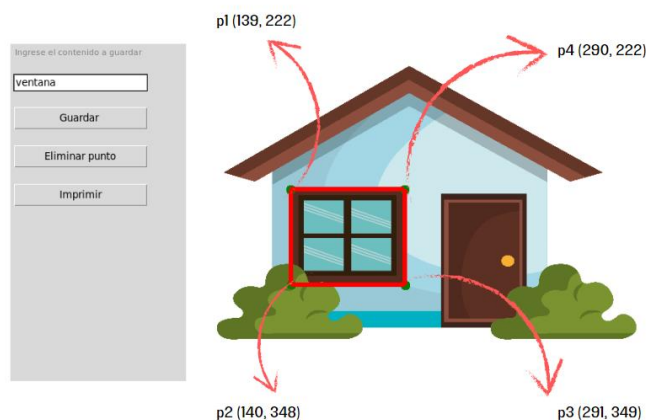
Definición del etiquetado. Cada zona de interés debe estar acompañada de una etiqueta que representa el nombre correspondiente de la composición de la imagen. Para ingresar esta etiqueta, se dispone de un espacio de texto en el que el usuario puede escribir el nombre deseado, como se muestra en la Figura 17.

Es importante destacar que la etiqueta debe ser ingresada para cada zona de interés individualmente. Por ejemplo, si la zona de interés representa una ventana, la etiqueta correspondiente deberá reflejar ese nombre o descripción específica.

Durante el etiquetado, se realiza una comprobación para verificar si ya existe un archivo .csv creado para almacenar los datos en la base de datos. Si el archivo ya existe, se continúa cargando la información en ese archivo existente. En caso contrario, si no se encuentra un archivo .csv previo, se crea uno nuevo para guardar los datos recopilados.

Figura 17

Etiquetado de la zona de interés

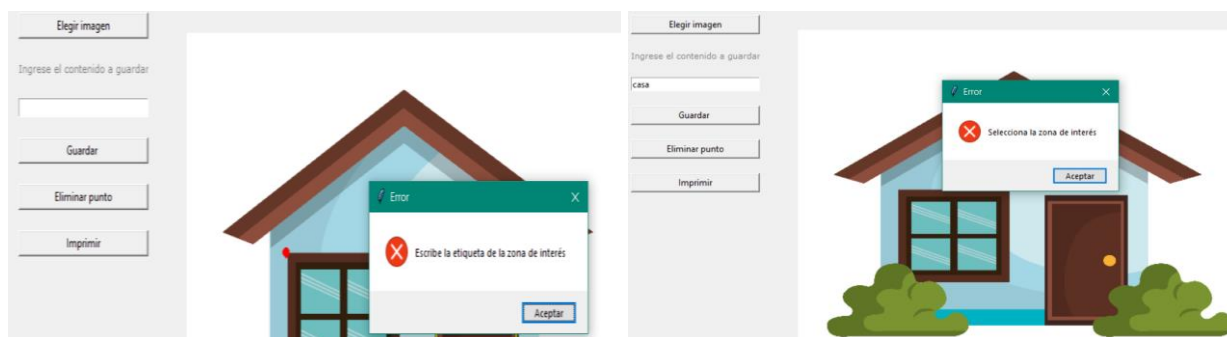


También incluye una verificación adicional para asegurarse de que existan puntos seleccionados y texto de etiqueta para cada zona de interés. Si se cumplen estas condiciones, se procede con el almacenamiento de la información.

Sin embargo, si no se cumplen estas condiciones, se activa un cuadro de mensaje informativo para alertar al usuario sobre la falta de puntos seleccionados o de etiqueta de zona de interés. Estos mensajes informativos se presentan de manera clara y comprensible para el usuario, brindando la información necesaria para corregir el problema. Un ejemplo de cómo se presentan estos mensajes informativos se ilustra en la Figura 18.

Figura 18

Ausencia de la etiqueta o zona de interés para el almacenamiento



Eliminación del último punto. La función de eliminación del último punto ha sido diseñada para permitir la remoción o eliminación del punto más reciente que se haya agregado, tanto en la lista de almacenamiento como en la representación gráfica del mismo.

Esta función se implementa con el propósito de corregir errores humanos que puedan ocurrir al seleccionar puntos mediante clics en la interfaz. En caso de que se haya realizado un clic de forma incorrecta o no deseada, la función de eliminación del último punto brinda la posibilidad de deshacer ese último punto agregado.

Al eliminar el último punto, también se eliminan las líneas que conectan ese punto con el punto anterior y con el primer punto, lo que ayuda a mantener la integridad visual del polígono formado por los puntos.

Almacenamiento de información en una base de datos. La función de almacenamiento de información en una base de datos se encarga de guardar los datos recopilados en un archivo .csv. Esta función se utiliza para almacenar de manera estructurada y organizada las variables relevantes para el etiquetado de la imagen.

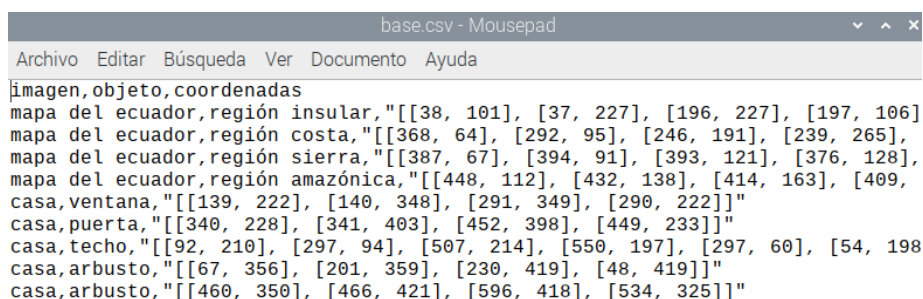
Los datos que se guardan en el archivo .csv incluyen las siguientes variables:

- "Nombre de la imagen": Esta variable representa el nombre o identificador de la imagen en la que se ha realizado el etiquetado de las zonas de interés.
- "Nombre de la etiqueta de la zona de interés": Esta variable corresponde al nombre o descripción asignada a la zona de interés específica dentro de la imagen.
- "Coordenadas obtenidas del polígono": Esta variable almacena las coordenadas que definen el polígono formado por los puntos de la zona de interés. Estas coordenadas permiten delimitar de manera precisa el área de interés en la imagen.

Se muestra en la Figura 19, el proceso de almacenamiento, tanto que con el ejemplo de etiquetado anteriormente visualizado se puede constatar en esta base de datos.

Figura 19

Almacenamiento de datos de etiquetado de la imagen



```

base.csv - Mousepad
Archivo Editar Búsqueda Ver Documento Ayuda
imagen,objeto,coordenadas
mapa del ecuador,región insular,"[[38, 101], [37, 227], [196, 227], [197, 106]
mapa del ecuador,región costa,"[[368, 64], [292, 95], [246, 191], [239, 265],
mapa del ecuador,región sierra,"[[387, 67], [394, 91], [393, 121], [376, 128],
mapa del ecuador,región amazónica,"[[448, 112], [432, 138], [414, 163], [409,
casa,ventana,"[[139, 222], [140, 348], [291, 349], [290, 222]]"
casa,puerta,"[[340, 228], [341, 403], [452, 398], [449, 233]]"
casa,techo,"[[92, 210], [297, 94], [507, 214], [550, 197], [297, 60], [54, 198
casa,arbusto,"[[67, 356], [201, 359], [230, 419], [48, 419]]"
casa,arbusto,"[[460, 350], [466, 421], [596, 418], [534, 325]]"

```

Generación de código QR. La función de generación de código QR se encarga de crear códigos QR a partir del nombre de la imagen, sin incluir la extensión del archivo. Estos códigos QR son útiles para identificar de manera única cada imagen preprocesada y facilitar el acceso a los datos almacenados correspondientes a esa imagen en la base de datos.

El código QR generado contiene información que permite asociar la imagen preprocesada con los datos almacenados en la base de datos. En la primera etapa del proceso, al ingresar la imagen preprocesada, se utiliza el código QR como un identificador para acceder a los datos específicos relacionados con el nombre de la imagen.

Como se muestra en la Figura 20, el código QR representa un identificador único que al ser escaneado muestra su contenido y en este caso se visualizará la palabra “casa” la cual da vinculación con los datos almacenados en la base de datos.

Figura 20

Identificador para acceder a la base de datos almacenado



Descarga de imagen en formato específico. La función de descarga de imagen en formato específico se encarga de generar un archivo PDF que contiene la imagen seleccionada, marcadores de ArUco y el código QR correspondiente. Este archivo PDF está diseñado en formato A4 y se crea con el propósito de que pueda ser impreso y utilizado por personas con discapacidad visual.

Al hacer clic en el botón "Imprimir", la función se activa y se genera el archivo PDF en una ubicación específica en el computador. Este material generado, como se muestra en la Figura 21, incluye la imagen preprocesada, los marcadores de ArUco que brindan orientación y referencia precisa, así como el código QR que facilita la identificación de la imagen y su asociación con los datos almacenados en la base de datos.

El archivo PDF obtenido a través de esta función permite obtener el material necesario para el uso por parte de la persona con discapacidad visual. Al imprimirlo, se obtiene una versión tangible de la imagen preprocesada

Figura 21

Material generado para la persona con discapacidad visual

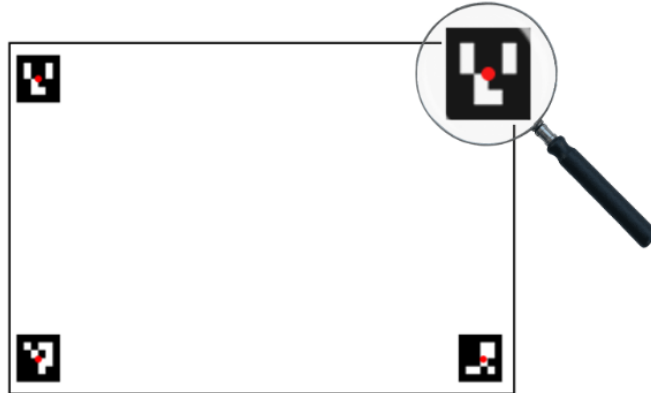


Corrección de orientación de la imagen

Las esquinas se reconocen con los marcadores ArUco que se encuentran dentro de un diccionario, en este caso, el diccionario es DICT_4X4_50, si bien, esta aplicación utiliza los cuatro primeros marcadores para identificar las esquinas de una hoja en formato A4. Los marcadores ArUco permite detectar las coordenadas de los centros de los marcadores como proyecta la Figura 22, para posterior con esa información y los IDs de los marcadores detectados, transformar la perspectiva según unas dimensiones preestablecidas.

Figura 22

Puntos céntricos de los marcadores de ArUco para estimar la orientación



La transformación de perspectiva se lo realiza con la biblioteca OpenCV, la que permite convertir una imagen desde una perspectiva tridimensional a bidimensional con una matriz de transformación con el propósito de capturar una región de toda la imagen, para trabajar con el mismo espacio y dimensión, en este cambio de perspectiva utiliza las cuatro coordenadas de las esquinas cómo también de la imagen original para devolver una imagen transformada como se visualiza en la Figura 23.

Figura 23

Transformación de perspectiva en tiempo real



Carga de información asociada a la imagen

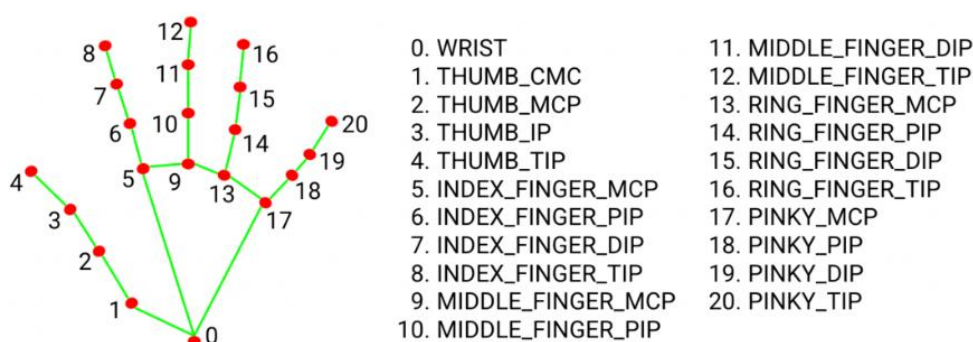
La información asociada a la imagen se adquiere mediante la decodificación del código QR y se obtiene la base de datos anteriormente ilustrada en la Figura 19, lo cual proporcionará el "Nombre de la imagen". Posteriormente, se realizará un escaneo de nombres semejantes en la base de datos .csv y se conocerá el "Nombre de la etiqueta de la zona de interés" y las "Coordenadas obtenidas del polígono" correspondientes.

Seguimiento de la posición dactilar

El seguimiento de la posición dactilar se lo realiza con la biblioteca MediaPipe, la cual posee un conjunto de modelos pre entrenado de aproximadamente 30000 imágenes de palmas reales para el reconocimiento, seguimiento y análisis de manos en tiempo real. Reconoce 21 coordenadas en una mano como se visualiza en la Figura 24 que pertenece a los nudillos (Google, 2021).

Figura 24

Nombres de los 21 puntos de referencia de una mano



Nota. Tomado de *Google for Developers: Hand landmarks detection guide*), por Google, 2023

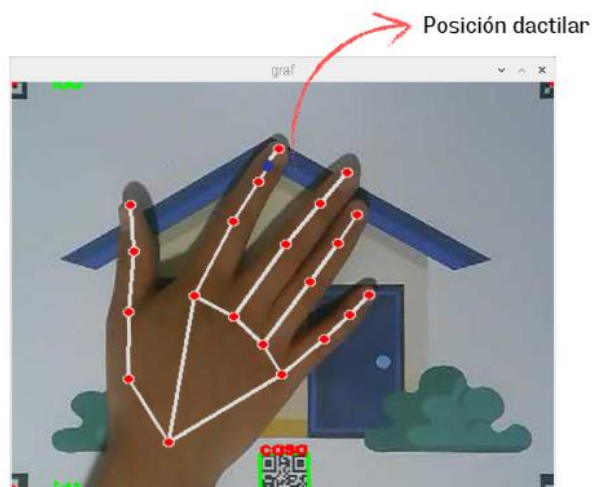
Detección de la posición dactilar en la posición de interés

La detección de la posición dactilar en el aplicativo se dibujan los puntos clave de la palma, dedos y la interconexión entre ellos, permitiendo la identificación del dedo índice la cual

se escogió el punto promedio entre el *INDEX_FINGER_TIP* y *INDEX_FINGER_DIP* con un círculo identificador como se visualiza en la Figura 25, la identificación se almacena en forma de coordenadas, la cual por el cambio de perspectiva también se redefinió a las dimensiones establecidas.

Figura 25

Posición del dedo índice en la zona de interés en tiempo real



Reproducción de mensaje asociado a la región de interés

La búsqueda y verificación se lleva a cabo mediante la carga de información de la imagen, donde se verifica si coincide con el texto decodificado del código QR en la base de datos almacenado. Además, se detectan las coordenadas a través del dedo índice, y se identifica el "Nombre de la etiqueta de la zona de interés". Esto es posible gracias al uso de la biblioteca shapely, que permite convertir las coordenadas en polígonos. Asimismo, la detección de la posición dactilar genera un punto de verificación con las coordenadas. Si el punto se encuentra dentro de algún polígono, se reproduce un audio de la etiqueta correspondiente utilizando la ayuda de gtts (*Google Text To Speech*) y playsound.

Capítulo IV

Pruebas y resultados

La evaluación del desempeño del prototipo que brinda accesibilidad a la composición de una imagen digital preprocesada mediante el seguimiento de la posición dactilar tiene como uno de los objetivos la validación, la eficiencia y efectividad del aplicativo en términos de usabilidad y accesibilidad en personas con discapacidad visual. Para lograr este propósito, se llevó a cabo un proceso de pruebas y observación con el fin de recopilar información relevante. Considerando que estas pruebas se realizan con seres humanos, este proyecto de investigación es no invasivo, salvaguardando la dignidad, derechos, bienestar y seguridad en los participantes en este espacio, que según el Reglamento sustitutivo para la aprobación y seguimiento de Comités de Ética de Investigación en Seres Humanos (CEISH) y Comités de Ética Asistenciales para la Salud (CEAS) del Ministerio de Salud Pública de la República del Ecuador menciona “Se consideran investigaciones sin riesgo, aquellos estudios en los que no se realiza ninguna modificación o intervención intencionada sobre variables biológicas, psicológicas o sociales de los sujetos participantes, y en los cuales no sea posible identificar a los mismos.” (Órgano de la República del Ecuador, 2022, p.19). Dado que la aplicación desarrolla retroalimentación auditiva de la composición de una imagen a través de un altavoz, los niveles de exposición de audio del dispositivo se mantienen dentro del rango normalizado por la OMS y la UIT. Según las pautas establecidas por estas organizaciones, para adultos el nivel es de 80 dBA (Sound Pressure Level - SPL), mientras que para usuarios sensibles es de 75 dBA SPL, garantizando así una experiencia libre de cualquier riesgo auditivo. (World Health Organization & International Telecommunication Union, 2019).

Metodología

La validación de la accesibilidad y usabilidad del prototipo en la composición de imágenes digitales preprocesadas mediante el seguimiento de la posición dactilar, se llevó a

cabo de manera experimental para evaluar la eficiencia y efectividad del aplicativo para personas con discapacidad visual.

Participantes

En la fase preliminar de pruebas, se extendió una invitación a una persona con discapacidad visual para participar. Esta etapa de pruebas se consideró crucial para identificar y corregir detalles, así como para perfeccionar el prototipo antes de continuar con las evaluaciones en un grupo más amplio.

En la fase total, se contó con la participación de diez personas con discapacidad visual afiliados a una asociación de ciclismo para personas ciegas pertenecientes a FEDEDIV (Federación Ecuatoriana De Deportes Para Personas Con Discapacidad Visual). Esta asociación facilitó la recopilación de información necesaria para validar la usabilidad y accesibilidad del prototipo.

Procedimiento de pruebas

La prueba se centró en la usabilidad con respecto a la accesibilidad de la composición de una imagen digital preprocesada mediante el seguimiento de la posición dactilar. Se buscó determinar si el prototipo creado ayuda a una persona con discapacidad visual al conocer la composición de una imagen y proporciona una experiencia satisfactoria al interactuar con las imágenes a través de la posición del dedo índice.

Antes de iniciar las pruebas, se realizó una sesión introductoria donde se explicó el propósito del estudio y se obtuvo el consentimiento de los participantes (Anexo A). Durante esta sesión, se les proporcionó una breve descripción de las funcionalidades del prototipo, se les dio la oportunidad de explorar a cada participante la aplicación con el objetivo de que conozcan el aplicativo y como este funciona.

La evaluación del prototipo se llevó a cabo mediante el uso de imágenes previamente seleccionadas (Anexo B) con el fin de verificar las zonas de interés correctamente identificadas. Durante esta evaluación, los participantes interactuaron con el prototipo donde se proporcionaron imágenes preprocesadas que se ubicaron una por una en la base del prototipo. Al colocar cada imagen, la cámara detectaba su nombre y recuperaba la información correspondiente de la base de datos creada. Esta información se reproducía a través de un altavoz, al comienzo de la colocación de la imagen inicia con la reproducción del nombre de la imagen. A medida que el participante exploraba la imagen, se reproducía su composición según la zona de interés en la que el dedo índice se encontraba.

Simultáneamente a la participación, se registraron los aciertos al identificar las zonas de interés como se observa en la Tabla 3 y Tabla 4. Se consideraba un acierto si la zona identificada coincidía con el área de la posición del dedo índice del participante. Por ejemplo, si una imagen contenía cuatro zonas de interés y el participante lograba identificar correctamente las cuatro, se registraban como zonas de interés correctamente reconocidas.

Figura 26

Evaluación de validación del prototipo con personas con discapacidad visual



La prueba contó con la participación de once personas que interactuaron con el prototipo entre la fase preliminar y la fase total.

Fase preliminar. Durante esta etapa preliminar, se llevó a cabo una evaluación con un único participante que presentaba discapacidad visual. Este participante tuvo la oportunidad de interactuar con un conjunto de diez imágenes. A partir de esta interacción, se obtuvieron los siguientes datos e información:

Tabla 3

Fase preliminar - Registro de zona de interés identificadas

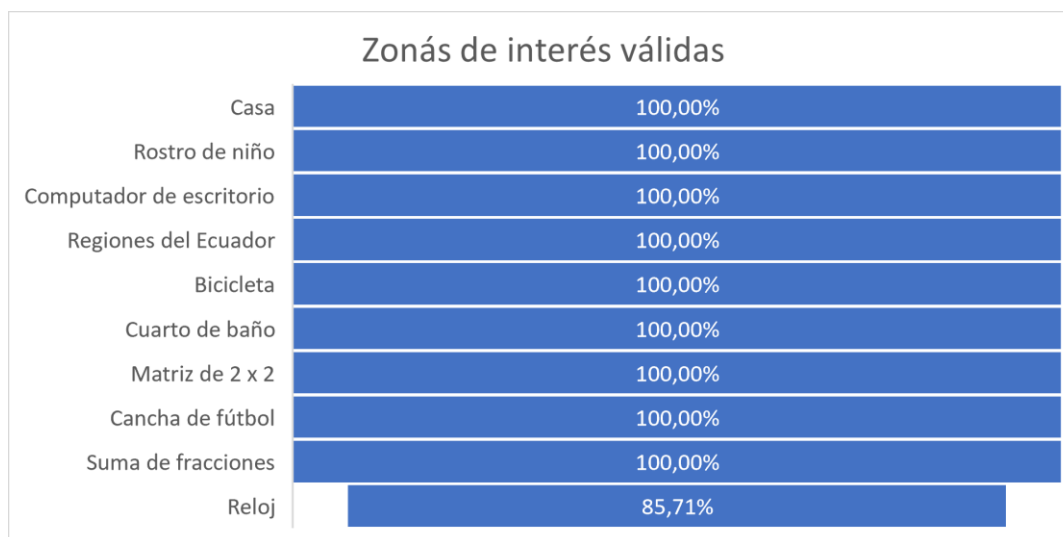
Participante	Imagen	Zonas de interés	Zonas de interés correctamente reconocidas
P1	Casa	4	4
	Rostro de niño	7	7
	Computador de escritorio	4	4
	Regiones del Ecuador	4	4
	Bicicleta	6	6
	Cuarto de baño	4	4
	Matriz de 2 x 2	8	8
	Cancha de fútbol	4	4
	Suma de fracciones	5	5
	Reloj	14	12

Basándonos en los datos recopilados y la información proporcionada por el participante, se puede concluir que en la fase preliminar se confirmó que el prototipo logró una precisión del

98,57% al identificar correctamente las zonas de interés. Sin embargo, se identificó la necesidad de abordar y corregir los aspectos observados para mejorar aún más el funcionamiento del sistema.

Figura 27

Zonas de interés válidas en la zona de interés correspondientes en la fase preliminar

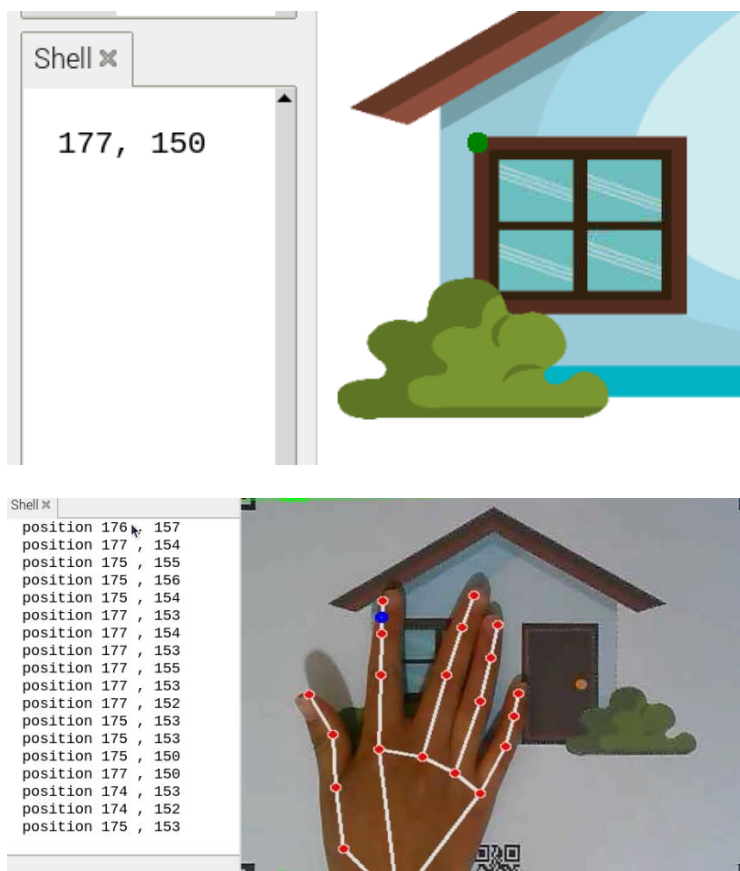


Los aspectos observados que requerían corrección para mejorar la precisión del prototipo fueron los siguientes: Se identificaron detalles que necesitaban ajustes para aumentar la precisión del prototipo. Una de las áreas modificadas se enfocó en la velocidad de detección de la posición dactilar, ya que se observó un cierto retraso en la actualización de la posición dactilar. Este retraso se debía a la creación de un buffer en la reproducción de mensajes, lo que resultó retención de mensajes previos en situaciones de movimientos rápidos de la posición dactilar. En otras palabras, cuando un cambio rápido de posición dactilar llevaba al usuario a una nueva área de interés, ocasionalmente se repetía el mensaje correspondiente a la zona de interés anterior lo que llevaba a generar confusión. Cabe destacar que esta confusión no se debía a errores en la limitación del área de la zona de interés, pues se verificó

con éxito su exactitud que la posición dactilar que estimaba un punto coincide con la aplicación y la que se detectaba como se visualiza en la Figura 28.

Figura 28

Concordancia de coordenadas entre lo seleccionado y en tiempo real



Fase total. En esta fase final, tras llevar a cabo las correcciones necesarias, los resultados se basan en la información recopilada de la participación de diez personas con discapacidad visual. Estos participantes tuvieron acceso a un conjunto de cinco imágenes específicas. Los datos obtenidos en esta etapa se presentan a continuación:

Tabla 4*Fase total - Registro de zona de interés identificadas*

Participante	Imagen	Zonas de interés	Zonas de interés correctamente reconocidas
P2	Casa	4	4
	Rostro de niño	7	7
	Computador de escritorio	4	4
	Regiones del Ecuador	4	4
	Bicicleta	6	6
P3	Casa	4	4
	Rostro de niño	7	7
	Computador de escritorio	4	3
	Regiones del Ecuador	4	4
	Bicicleta	6	6
P4	Casa	4	4
	Rostro de niño	7	7
	Computador de escritorio	4	4
	Regiones del Ecuador	4	4
	Bicicleta	6	6
P5	Casa	4	4
	Rostro de niño	7	7
	Computador de escritorio	4	4
	Regiones del Ecuador	4	4
	Bicicleta	6	6

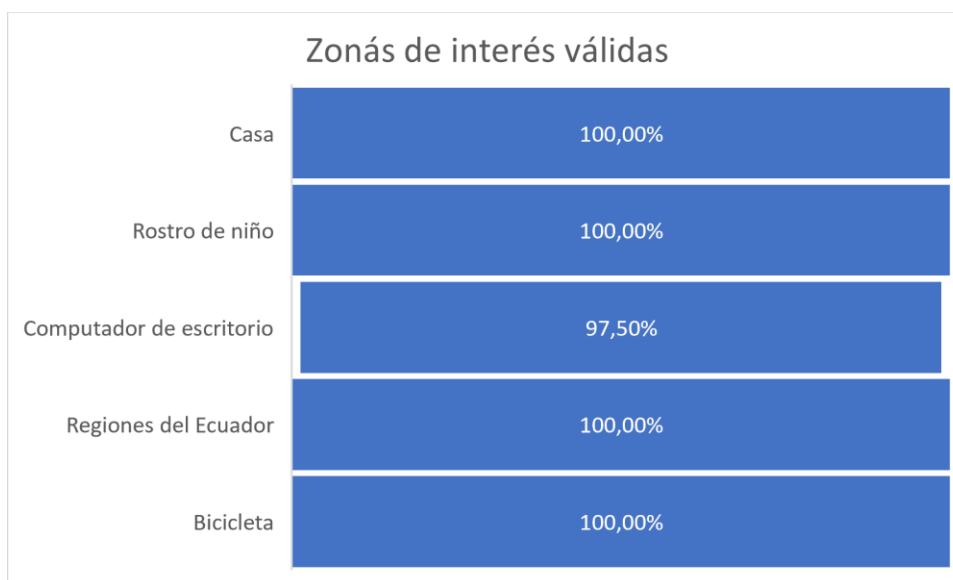
Participante	Imagen	Zonas de interés	Zonas de interés correctamente reconocidas
P6	Casa	4	4
	Rostro de niño	7	7
	Computador de escritorio	4	4
	Regiones del Ecuador	4	4
	Bicicleta	6	6
P7	Casa	4	4
	Rostro de niño	7	7
	Computador de escritorio	4	4
	Regiones del Ecuador	4	4
	Bicicleta	6	6
P8	Casa	4	4
	Rostro de niño	7	7
	Computador de escritorio	4	4
	Regiones del Ecuador	4	4
	Bicicleta	6	6
P9	Casa	4	4
	Rostro de niño	7	7
	Computador de escritorio	4	4
	Regiones del Ecuador	4	4
	Bicicleta	6	6
P10	Casa	4	4
	Rostro de niño	7	7

Participante	Imagen	Zonas de interés	Zonas de interés correctamente reconocidas
P11	Computador de escritorio	4	4
	Regiones del Ecuador	4	4
	Bicicleta	6	6
	Casa	4	4
	Rostro de niño	7	7
	Computador de escritorio	4	4
	Regiones del Ecuador	4	4
	Bicicleta	6	6

Referenciándose en los datos recopilados durante esta etapa se confirma que el prototipo logra identificar de manera precisa las zonas de interés con una precisión del 99,5%. Esta notable precisión otorga validez a la accesibilidad de la composición de imágenes, cumpliendo con éxito los criterios establecidos en la prueba. Además, estos resultados demuestran que el prototipo ha alcanzado su objetivo de proporcionar una accesibilidad a la composición de imágenes para personas con discapacidad visual.

Figura 29

Zonas de interés válidas en la zona de interés correspondientes en la fase última y definitiva



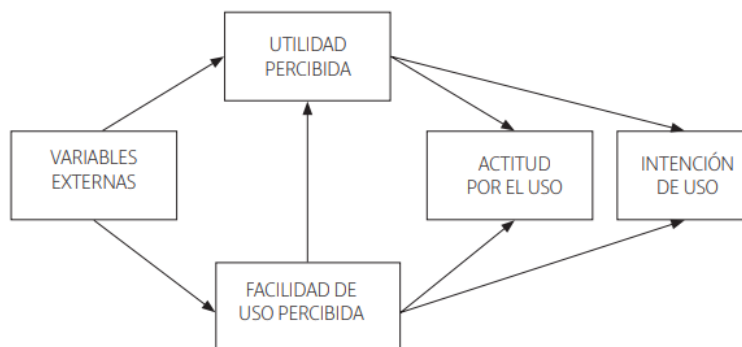
Las correcciones realizadas, en particular la mejora en la velocidad de detección dactilar, han generado mejoras significativas tanto en la precisión como en la experiencia del usuario permitiendo reforzar la efectividad y utilidad del prototipo.

Al concluir cada participación, se administró una encuesta diseñada para evaluar la aceptación de la tecnología según el criterio de las personas con discapacidad visual. La encuesta incluyó preguntas cerradas que permitieron a los participantes expresar su opinión en una escala del 1 al 5, donde el valor 5 representaba la ponderación más alta. Estas preguntas se adhirieron al enfoque propuesto de Davis (1989) que en la investigación de Cabero-Almenara & Pérez Díez De Los Ríos (2018) describe el Modelo de Aceptación de Tecnología (Technology Acceptance Model - TAM) que determina las respuestas de los usuarios hacia la tecnología en función de dos variables principales: la utilidad percibida y la facilidad de uso. Estas variables influyen en la actitud hacia el uso de la tecnología y, a su vez, en la intención de uso, pudiendo generar una respuesta favorable o desfavorable hacia su adopción e intención de usarla. Por tanto, es útil adicionar una encuesta de usabilidad está basada en la

formulación inicial del Modelo de Aceptación de Tecnología (TAM) cómo se puede ilustrar en la Figura 30.

Figura 30

Formulación inicial de TAM



Las preguntas establecidas para la encuesta fueron las siguientes:

1. Utilidad percibida

¿Has encontrado que la aplicación te permite reconocer las zonas de interés y entender de qué imagen se realiza la descripción de la composición?

2. Facilidad de uso percibida

¿Cómo calificarías la facilidad de uso general de la aplicación?

3. Actitud por el uso

¿Consideras que la aplicación proporciona información clara y comprensible a través de medios no visuales, como el sonido o la retroalimentación táctil?

4. Intención de uso

Considerando tu experiencia con la aplicación hasta el momento, ¿qué tan probable es que sigas utilizando esta tecnología en el futuro para llevar a cabo tus tareas o actividades habituales?

Posterior al realizar la encuesta a los participantes con discapacidad visual se obtuvieron los siguientes resultados:

Tabla 5

Registro de puntuación de la encuesta para la aceptación del prototipo

Participante	Utilidad percibida	Facilidad de uso percibida	Actitud por el uso	Intención de uso
P1	5	4	5	5
P2	5	4	5	5
P3	4	4	5	5
P4	5	5	5	3
P5	5	5	5	3
P6	3	4	4	4
P7	3	5	3	5
P8	5	4	4	5
P9	5	5	5	5

Participante	Utilidad percibida	Facilidad de uso percibida	Actitud por el uso	Intención de uso
P10	5	5	5	5
P11	5	4	5	5

1. Utilidad Percibida:

- 90.9% de los participantes calificaron la aplicación con una utilidad percibida de 5.
- 9.1% de los participantes calificaron la aplicación con una utilidad percibida de 4.

2. Facilidad de Uso Percibida:

- 81.8% de los participantes calificaron la facilidad de uso percibida con un 5.
- 18.2% de los participantes calificaron la facilidad de uso percibida con un 4.

3. Actitud por el Uso:

- 90.9% de los participantes calificaron la actitud por el uso con un 5.
- 9.1% de los participantes calificaron la actitud por el uso con un 4.

4. Intención de Uso:

- 54.5% de los participantes indicaron una intención de uso futura con un 5.
- 36.4% de los participantes indicaron una intención de uso futura con un 4.
- 9.1% de los participantes indicaron una intención de uso futura con un 3.

Este análisis resalta que la mayoría de los participantes tuvieron una percepción muy positiva en cuanto a la utilidad, facilidad de uso y actitud por el uso de la aplicación. Además, más de la mitad expresaron una alta intención de seguir utilizando la tecnología en el futuro para sus tareas y actividades de aprendizaje tales como cursos y entrenamientos educativos.

La combinación de datos cuantitativos y cualitativos revela que el prototipo logró su objetivo de proporcionar accesibilidad en la composición de imágenes para personas con discapacidad visual. Además, los altos índices de aceptación en la encuesta TAM respaldan la relevancia y utilidad de la tecnología en el contexto de personas con discapacidad visual. Esta aceptación resuena como un respaldo sólido para la trascendencia de la tecnología en la vida de las personas con discapacidad visual.

Capítulo V

Conclusiones y trabajos futuros

Conclusiones

Se ha logrado desarrollar un prototipo innovador que brinda accesibilidad a la composición de imágenes digitales preprocesadas mediante el seguimiento de la posición dactilar en tiempo real, dirigido específicamente a personas con discapacidad visual. A través de la integración de tecnologías como la inteligencia artificial, el procesamiento de imágenes y la retroalimentación auditiva, se ha creado una solución que aborda de manera efectiva las limitaciones en la percepción visual de este grupo de usuarios.

Para determinar el algoritmo óptimo para la detección dactilar se realizó una investigación exhaustiva la que involucró la biblioteca MediaPipe, aprovechando las capacidades de esta biblioteca, se adaptó su funcionalidad para satisfacer las necesidades precisas del prototipo, específicamente en relación con la detección en tiempo real de la posición y ubicación del dedo índice.

Este prototipo utilizó una tarjeta Raspberry Pi como microcomputadora principal, elegida debido a su bajo costo y dimensiones compactas. Se optó por trabajar directamente con Python 3, aprovechando su amplio acceso a valiosas bibliotecas de código abierto, como MediaPipe, OpenCV, GTTS y Tkinter, que fueron de utilidad para la implementación de la solución.

Se implementó una interfaz gráfica que posibilitó el preprocesamiento de imágenes con el propósito de limitar y etiquetar las zonas de interés. Esta aplicación ayuda a generar una base de datos donde se almacenó información esencial asociada a cada imagen, incluyendo su "Nombre", la "Etiqueta de la zona de interés" y las "Coordenadas del área de la zona de interés". Además, la aplicación facilita la generación de una forma de desarrollar el material didáctico de las imágenes, esta imagen contiene marcadores ArUco en las esquinas de una

hoja en A4 para orientación y cambio de perspectiva, así como un Código QR para identificación y acceso a la base de datos mediante el nombre de la imagen guardada en un código QR y la propia imagen en cuestión.

Se desarrolló la aplicación que permitió la detección dactilar sobre la imagen y proporcionó retroalimentación auditiva correspondiente a la zona de interés identificada.

La reproducción de audio se realiza mediante la obtención de información de la base de datos, que es decodificada a partir del código QR. Además, los marcadores ArUco, ubicados en las esquinas, posibilitaron la redefinición de las coordenadas de posición dactilar en función de la nueva perspectiva, asegurando una precisión adecuada en la información proporcionada.

En la evaluación con personas discapacitadas visuales, el prototipo logró una precisión del 99.5% al identificar zonas de interés en imágenes, validando su accesibilidad. En términos de usabilidad según el modelo TAM, la mayoría calificó positivamente la utilidad, facilidad de uso y actitud, mostrando aceptación. Además, más del 90% expresó intención de usar el prototipo en el futuro para actividades de aprendizaje, indicando alta aceptación.

Trabajos futuros

En la búsqueda de mejoras y expansiones futuras, se plantea como continuidad a este proyecto de titulación la implementación de la personalización de la velocidad de reproducción audible, ajustable según las preferencias del usuario, este enfoque permitiría a los usuarios adaptar la experiencia a sus propias necesidades y preferencias individuales, lo que podría resultar en una experiencia más fluida y efectiva para las personas con discapacidad visual.

Una siguiente necesidad evidente es la implementación de una aplicación móvil que pueda ser ejecutada en teléfono inteligente y que ofrezca las mismas funcionalidades que la descripción de la composición de imágenes a través de la posición dactilar en tiempo real. Esta evolución se justifica por la necesidad de aumentar la portabilidad y accesibilidad del prototipo,

brindándoles una mayor independencia y empoderamiento en su vida cotidiana para superar las restricciones actuales asociadas a su robustez y limitaciones de uso en diversos entornos.

Adicional a ello, se pudo rescatar necesidades en las personas con discapacidad visual donde pretenden escuchar la composición de una imagen en un nivel más detallado, que incluya características específicas de cada zona de interés, como forma, color y dimensiones. Esta extensión podría contribuir significativamente a un mayor nivel de entendimiento cognitivo.

Un posible trabajo futuro sería ampliar la base de datos de este prototipo, diseñado para brindar accesibilidad a la composición de imágenes según el seguimiento dactilar. Agregar información sobre billetes y monedas que beneficiaría a personas con discapacidad visual, mejorando su comprensión y manejo del dinero.

Bibliografía

Aguirre, T., & Olmedo, R. (2016). Sistema 3D de Integración de Datos de Ensayos Aerodinámicos en Túnel de Viento. *Revista Tecnología y Ciencia, Universidad Tecnológica Nacional*. https://www.researchgate.net/profile/Miguel-Angel-Aguirre/publication/315692848_Sistema_3D_de_Integracion_de_Datos_de_Ensayos_Aerodinamicos_en_Tunel_de_Viento/links/58dbe6e0458515152b2ec149/Sistema-3D-de-Integracion-de-Datos-de-Ensayos-Aerodinamicos-en-Tunel-de-Viento.pdf

Álvarez Vega, M., Quirós Mora, L. M., & Cortés Badilla, M. V. (2020). Inteligencia artificial y aprendizaje automático en medicina. *Revista Medica Sinergia*, 5(8), e557. <https://doi.org/10.31434/rms.v5i8.557>

Amos, D. (2020). *Python GUI Programming With Tkinter*.

Anzola, J. P. A. (2022). *La Robótica y Automatización Industrial: Una perspectiva desde el aprendizaje automático*.

Cabero-Almenara, J., & Pérez Díez De Los Ríos, J. L. (2018). Validación del modelo TAM de adopción de la Realidad Aumentada mediante ecuaciones estructurales. *Estudios sobre Educación*, 34, 129-153. <https://doi.org/10.15581/004.34.129-153>

CONADIS. (2022, enero). *Estadísticas de Discapacidad*. Tableau Software. https://public.tableau.com/views/Discapacidad/Inicio?:embed=y&:showVizHome=no&:loadOrderID=0&:display_count=yes&:showTabs=y

Davis, F. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13, 319. <https://doi.org/10.2307/249008>

Dewi, C., & Juli Christanto, H. (2022). Combination of Deep Cross-Stage Partial Network and Spatial Pyramid Pooling for Automatic Hand Detection. *Big Data and Cognitive Computing*, 6(3), Article 3. <https://doi.org/10.3390/bdcc6030085>

Domínguez, T. M. (2021). *Visión artificial: Aplicaciones prácticas con OpenCV - Python* (Marcombo).

Gonzalez, R., & Woods, R. (2002). *Digital Image Processing (Second Edition)*. Prentice-Hall.
https://www.researchgate.net/publication/333856607_Digital_Image_Processing_Second_Edition

Google. (2021). *Hand landmarks detection guide | MediaPipe*. Google for Developers.
https://developers.google.com/mediapipe/solutions/vision/hand_landmarker

Intel Corporation. (s. f.). *Redes neuronales convolucionales (CNN) y aprendizaje profundo*. Intel. Recuperado 11 de julio de 2023, de
<https://www.intel.com/content/www/es/es/internet-of-things/computer-vision/convolutional-neural-networks.html>

Jácome, J., & Viracocha, E. (2019). *Desarrollo de una aplicación para detección de patrones en imágenes mediante el uso de aprendizaje profundo*. UNIVERSIDAD CENTRAL DEL ECUADOR.

Jara, N., Abad, J. L., Navarro, J., & Page, Á. (2019, julio 18). Validación de los marcadores ArUco para el análisis de movimientos humanos. *Libro de Actas - 11 Simposio CEA de Bioingeniería*. 11 Simposio CEA de Bioingeniería.
<https://doi.org/10.4995/CEABioIng.2019.10029>

Khanagar, S. B., Al-ehaideb, A., Maganur, P. C., Vishwanathaiah, S., Patil, S., Baeshen, H. A., Sarode, S. C., & Bhandi, S. (2021). Developments, application, and performance of artificial intelligence in dentistry – A systematic review. *Journal of Dental Sciences*, 16(1), 508-522. <https://doi.org/10.1016/j.jds.2020.06.019>

Kuriakose, B., Shrestha, R., & Sandnes, F. E. (2022). Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review. *IETE Technical Review*, 39(1). <https://doi.org/10.1080/02564602.2020.1819893>

Lewis, S. (2019). Intra- and Interpersonal Factors in Adolescence Predicting Loneliness among Young Adults with Visual Impairments. *Journal of Visual Impairment & Blindness*, 113(1), 5. <https://doi.org/10.1177/0145482X18825184>

López de Paz, G. J. (2012). *Diseño de un programa de ortorectificación y georeferenciación de imágenes aéreas aplicadas a campos de caña de azúcar*. https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/5356/LOPEZ_GARY_DISE%C3%91O_PROGRAMA_ORTORECTIFICACION_GEOREFERENCIACION_IMAGENES_AERAS_CAMPOS_CA%C3%91A_AZUCAR.pdf?sequence=1

McGookin, D., Robertson, E., & Brewster, S. (2010). Clutching at straws: Using tangible interaction to provide non-visual access to graphs. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1715-1724. <https://doi.org/10.1145/1753326.1753583>

Nieto, R. L. (2012). *Aplicación de las tecnologías de la información y las comunicaciones en la vida diaria de las personas con discapacidad*. Universidade da Coruña, Servizo de Publicacións ; Fundación Orange.

Órgano de la República del Ecuador. (2022). *Reglamento sustitutivo del Reglamento para la aprobación y seguimiento de Comités de Ética de Investigación en Seres Humanos (CEISH) y Comités de Ética Asistenciales para la Salud (CEAS) del Ministerio de Salud Pública*

(N° 00005-2022; Quinto Suplemento N° 118, p. 92).

http://esacc.corteconstitucional.gob.ec/storage/api/v1/10_DWL_FL/eyJjYXJwZXRhIjoicm8iLCJ1dWkljoiNTJhOTFmYzgtNjQ0Yy00NTdlLWJhYzktZTU5MTY4M2I4ZmE2LnBkZiJ9

Python. (2023). *La biblioteca estándar de Python*. Python documentation.

<https://docs.python.org/3/library/index.html>

Raspberry Pi. (2019). *Raspberry Pi 4 Model B specifications*. Raspberry Pi.

<https://www.raspberrypi.com/products/raspberry-pi-4-model-b/specifications/>

Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G.-Z. (2017). Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1), 4-21. <https://doi.org/10.1109/JBHI.2016.2636665>

Richardson, M., & Wallace, S. P. (2016). *Getting started with Raspberry Pi: Getting to know the inexpensive ARM-powered Linux computer* (Third edition). Maker Media.

Vásquez, J. P. (2014). Red neuronal feedforward como estimador de patrones de corrientes en el interior del Puerto de Manzanillo sujeto a la acción de tsunamis. 406.

<https://www.imt.mx/archivos/Publicaciones/PublicacionTecnica/pt406.pdf>

World Health Organization & International Telecommunication Union. (2019). *Toolkit for safe listening devices and systems*. World Health Organization.

<https://apps.who.int/iris/handle/10665/280086>

Zador, A. M. (2019). A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications*, 10(1), 3770. <https://doi.org/10.1038/s41467-019-11786-6>

Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., & Grundmann, M. (2020). *MediaPipe Hands: On-device Real-time Hand Tracking* (arXiv:2006.10214). arXiv. <http://arxiv.org/abs/2006.10214>