

Título del Proyecto de Investigación
al que corresponde el Reporte Técnico:

Sistema auxiliar en hogares para personas invidentes basado en IoT

Tipo de financiamiento

Sin financiamiento

Fecha de Inicio: 15/11/2021
Fecha de Término: 25/11/2021

Tipo de Reporte

Parcial

Final

Autor (es) del reporte técnico:

Dr. Luis Carlos Méndez González
Dr. Luis Alberto Rodríguez Picón
Dr. Iván Juan Carlos Pérez Olguín
Dr. Luis Asunción Pérez Domínguez
Mtro. Abel Eduardo Quezada Carreón
Dr. David Luviano Cruz
Mtra. Luz Angelica García Villalba
Mtra. Fabiola Lom Monárrez
Mtro. Héctor Loya Caraveo.

Sistema auxiliar en hogares para personas invidentes basado en IoT

Resumen del reporte técnico en español

Se desarrolló una aplicación para dispositivos móviles la cual combina el cómputo en la nube con un sistema de visión artificial que, mediante una interfaz diseñada para reaccionar a comandos de voz, es capaz de identificar objetos dentro de un entorno cerrado. El producto final es en una aplicación para el sistema operativo Android que permite a usuarios invidentes detectar elementos mediante la cámara un dispositivo móvil, las imágenes capturadas se comparan con una base de datos, de manera que la interfaz delimita la posición del objeto detectado además de generar una etiqueta que indica la clase a la que este pertenece. Los resultados obtenidos permiten concluir que, el sistema es capaz de realizar con éxito consultas por medio de comandos de voz para identificar elementos a partir de la información almacenada en la base de datos con un porcentaje de confianza satisfactorio y un tiempo de consulta apropiado.

Resumen del reporte técnico en inglés:

An application for mobile devices was developed that combines cloud computing with an artificial vision system that can identify objects within a closed environment through an interface designed to react to voice commands. The final product is an application for the Android operating system that allows blind users to detect elements through a mobile device's camera. The captured images are compared with a database so that the interface delimits the detected object's position and generates a label that indicates the class to which it belongs. The results obtained allow us to conclude that the system can successfully carry out consultations through voice commands to identify elements from the information stored in the database with a reasonable confidence percentage and an appropriate consultation time.

Palabras clave:

Tecnologías asistenciales, Aplicación móvil, Visión artificial, Discapacidad Visual, Localización en interiores

Usuarios potenciales

Comunidad Científica Internacional y la empresa solicitante del reporte.

Reconocimientos

Agradecimientos a Alan Iván Hernández, estudiante de la carrera de Ing. en Mecatronica por su ayuda en el desarrollo de este proyecto.

1. INTRODUCCIÓN

El concepto de Internet de las cosas IoT (Por sus siglas en ingles Internet of Things) es algo que ha estado presente en muchos de los aspectos de la vida cotidiana de las personas por algo más de tres décadas, prácticamente desde el inicio mismo de la Internet, sin embargo, en la actualidad ha tomado una mayor relevancia con el mercado creciente de las aplicaciones para dispositivos inteligentes. La parte del “Internet” es lo que permite a múltiples dispositivos electrónicos estar interconectados entre ellos y de esta forma tener la posibilidad de acceder a una gran cantidad de información, mientras que el aspecto de “las cosas” hace referencia a todo aquel elemento físico con el que una persona tiene contacto en su vida diaria, ya sea su computadora, su automóvil, o incluso su propio teléfono móvil. En general, todos estos elementos comparten la característica de contar con sensores o circuitos integrados que permiten el manejo de datos, ya sea con otros dispositivos cercanos o mediante una conexión a Internet. [1]

En años recientes, uno de los mayores retos para la tecnología, consiste en poder superar las brechas que son ocasionadas a causa de una discapacidad, ya sea social, laboral o incluso emocional. En el caso de una discapacidad visual, el principal problema al que se enfrentan las personas es tener que depender de forma permanente de una herramienta que les permita desarrollar sus actividades diarias con normalidad, como lo suelen ser los

bastones, perros lazarillos y sistemas de señalamiento para personas invidentes como por ejemplo el Braille.

De la misma forma, el mercado actual de teléfonos inteligentes, así como el de las aplicaciones para cada sistema operativo, no suelen estar enfocadas a personas que padezcan algún problema visual, por lo que las herramientas que se pueden encontrar suelen tener una gran cantidad de limitaciones en cuanto a las funciones que ofrecen, además de que comúnmente suelen ser servicios de paga o de acceso limitado especialmente a la región en la que se desarrolló. [2]

La visión artificial tiene como objetivo el reproducir el sentido de la vista de las personas e implementarlo en una máquina de tal forma que, a partir de la información obtenida de una imagen, el algoritmo del sistema decida qué acción se deberá de realizar a continuación. Por la parte del hardware de un sistema de visión se suelen utilizar sensores que sean capaces de detectar, procesar y analizar imágenes en tiempo real como lo es en el caso de una cámara digital; mientras que para el software del sistema se necesita de un algoritmo que sea capaz de analizar y procesar la información proveniente de una imagen, para de esta forma poder crear una descripción detallada del entorno. [3]

2. PLANTEAMIENTO

De acuerdo con el INEGI, a través de su Encuesta Nacional de Dinámica Demográfica, en 2018 había alrededor de 7.8 millones de personas con alguna discapacidad en México, de las cuales, aproximadamente el 39% corresponde a individuos con algún problema visual que no puede ser corregido con el uso de anteojos [4]. Este grupo de personas puede presentar problemas visuales hasta el punto de padecer lo que se conoce como ceguera permanente; problema que afecta su calidad de vida hasta el punto de limitar su movilidad y afectar la forma en la que realizan sus actividades diarias, particularmente en el hogar. Para tratar esta cuestión, se utilizan herramientas la escritura braille [5] dispositivos hápticos [6] y sistemas de guía por voz [7], las cuales pueden facilitar sus actividades diarias. Incluso se han desarrollado dispositivos electrónicos que adaptan herramientas de navegación tradicionales, como los bastones [8]. Sin embargo, se pueden presentar situaciones, como la necesidad de buscar objetos o

de contar con sistemas de apoyo como guías, que permitan a este tipo de personas realizar sus actividades diarias con relativa facilidad.

2.1 Antecedentes

La visión por computadora o artificial abarca una serie de técnicas que implica la utilización herramientas de software para el diseño de algoritmos que involucren la aplicación de técnicas de aprendizaje maquina (como redes neuronales) con el fin de entrenar sistemas con información almacenada en bases de datos para que puedan reconocer patrones en ciertas características, como el color y la forma, de determinados objetos [1]. Estas técnicas, en conjunto con el procesamiento de imágenes y el reconocimiento facial y de objetos, tratan de emular el sentido de la visión humana. [2].En el área de visión artificial, con enfoque en el procesamiento de imágenes, se han diseñado técnicas de reconocimiento de colores basadas en el estándar RGB, con integraciones de interfaces Text-To-Speech, en sistemas de identificación de colores [3]; se han adaptado herramientas de accesibilidad para personas invidentes por medio de sistemas que utilizan retorno de audio para informar al usuario de la forma o el color de un determinado objeto [4]; se han desarrollado sistemas de reconocimiento y localización de objetos en el entorno de una casa con soporte para internet de las cosas y servicios en la nube [5]; se han implementado técnicas de análisis de patrones para retroalimentar sistemas de asistencia inteligentes para monitorear y detectar particularidades en las rutinas diarias de un usuario [6]; y se han creado dispositivos que integran sistemas de reconocimiento facial y detección de objetos para proporcionar herramientas de asistencia visual para personas invidentes [7].El computo en la nube consiste en proporcionar acceso a recursos escalables y virtualizados por medio de un esquema de servicio a través del internet [8] y el internet de las cosas (IoT)consiste en un sistema interconectado de dispositivos de cómputo, maquinas, dispositivos inteligentes y objetos que cuentan con la habilidad de transferir información a través de una red sin requerir interacciones entre hombre y maquina [9]. En el mundo actual, debido a la abundancia de dispositivos inteligentes, se facilita la implementación de esta

tecnología, que podría ofrecer ventajas como obtener información de cada dispositivo y almacenarla en la nube, para luego implementar un sistema basado en inteligencia artificial que utilice tal información para el aprendizaje maquina [10]. En el desarrollo del campo del cómputo en la nube y del IoT, estas tecnologías se han implementado en proyectos de domótica para proponer sistemas mejorados enfocados en el bajo consumo y una mejor conectividad entre dispositivos con servicios en la nube [11]; se han establecido sistemas de monitoreo y control para lámparas inteligentes [12]; y se ha desarrollado un análisis de la implementación de internet de las cosas y visión por computadora para monitorear el comportamiento de clientes y usuarios en entornos determinados, como negocios, para entrenar sistemas inteligentes que brinden asistencia y obtengan información de sus intereses [13]. Debido a la versatilidad del IoT, también se han realizado implementaciones enfocadas en las personas invidentes, como sistemas de puertas inteligentes con conectividad con servicios en la nube y otros dispositivos inteligentes [14]; se ha logrado la integración de aplicaciones móviles para facilitar la navegación de este tipo de usuarios en entornos con infraestructura IoT [15]; y la implementación de computación de borde para reducir la carga del ancho de banda y mejorar la respuesta de sistemas basados en IoT [16]. La accesibilidad consiste en proveer de medios que permitan ajustar un entorno a las necesidades y preferencias del usuario [17]; es decir, se basa en la implementación de sistemas, dispositivos y métodos de asistencia para facilitar las actividades diarias de las personas, con enfoque especial en las que cuentan con alguna discapacidad. De forma particular, las persona.

2.2 Marco teórico

El marco teórico de esta investigación está fundamentado en los siguientes temas

a. Visión por computadora

La visión por computadora consiste en emplear una computadora para, a través de un cámara o un dispositivo que actúe como un sensor de imágenes, obtener información relevante de un conjunto de imágenes o videos de objetos del mundo físico [2]. El enfoque principal de esta disciplina consiste en la implementación de herramientas computacionales para emular la función biológica de la visión; es decir, conseguir que una computadora pueda emular la capacidad de los seres vivos (particularmente, del ser humano) de obtener información del entorno, a través de la visión, y actuar con respecto a esta.

Dentro del campo de la visión por computadora se distinguen dos niveles de procesamiento: la visión preliminar y el análisis de la escena. La primera parte se considera de bajo nivel, ya que consiste en los primeros pasos de la visión para identificar características de ciertos objetos, como forma, color, posición, tamaño, entre otros; las cuales son utilizadas para realizar una segmentación progresiva de la escena capturada por el sistema de visión; es decir, descomponer una imagen en regiones específicas, que pueden coincidir con cambios notorios en las características detectadas (como en el caso del color y la forma, si se trata de un conjunto de frutas). La segunda parte del proceso de procesamiento implica un proceso de alto nivel en que se realiza una abstracción de las características obtenidas de la visión preliminar [5]. De acuerdo con el enfoque de la visión, se puede realizar tareas de análisis de formas, reconocimiento o localización de objetos; lo cual puede ser aplicado en procesos más complejos como el reconocimiento facial. El producto de la visión por computadora se genera en esta segunda etapa y consiste en la obtención de descripciones de alto nivel. En la figura 1 se muestra un panorama amplio sobre los conceptos de visión por computadora.

b. Algoritmos de reconocimiento de objetos

Una de las etapas más importantes de este proyecto es el algoritmo de reconocimiento de objetos. Ya que este sistema es la parte medular que realizara el sistema de decisión sobre el objeto que va a determinar cuál es. Dentro del procesamiento

de imágenes, detectar bordes de un objeto resulta ser vital y de gran utilidad, ya que facilitar las tareas de reconocimiento, segmentación y procesamiento.

Al hablar de bordes dentro de una imagen digitalizada, esta puede ser definida en diferentes niveles de escala, las cuales pueden estar a nivel de grises o de alguna escala de color. Con esta información, los objetos pueden ser identificados por cualquier algoritmo de Aprendizaje máquina. “Cuando se aborda un problema de reconocimiento de objetos usualmente se sigue un proceso básico de cinco pasos que son: [6]”.

- **Adquisición de Imagen:** En esta etapa se realiza la captura del objeto que se encuentre fijo o en movimiento para que posteriormente este puede ser digitalizado, almacenado para ser procesado por alguna inteligencia artificial.
- **Etapa de Preprocesamiento:** Aquí la imagen adquirida es contrastada con diferentes técnicas de mejoramiento de imagen como filtros, mejorar el contraste y la reducción del ruido.
- **Segmentación de la Imagen:** Una vez realizado el preprocesamiento, la imagen es dividida en regiones para aislar los objetos de interés, en esta etapa se extraen los metadatos de la imagen con la cual se pueden describir los rasgos de forma de vectores numéricos.
- **Clasificación de Imagen:** Las imágenes segmentadas son asignadas en clases o categorías, las cuales se basan en los rasgos de los metadatos obtenidos en la etapa de preprocesamiento

“Para el reconocimiento de formas se ha empleado un reconocedor de formas geométricas el cual emplea internamente la detección de bordes. Existen multitud de técnicas para la detección de bordes y a continuación se presentan uno de los más comunes” [4]

Algoritmo YOLO:

YOLO es un sistema de Código abierto en la cual se puede realizar la detección de objetos en tiempo real, para ello este hace uso de una red neuronal única, esta red se

establece de forma convolucional. El algoritmo determina por medio de regiones la probabilidad y las identifica en cajas para realizar una ponderación a partir de matemática estadística especializada. La ventaja de este algoritmo es a medida que van obteniendo más regiones y más segmentos va aprendiendo con lo cual durante cada interacción va reduciendo el error transitorio en cada conjunto de imágenes almacenadas. En la figura 2, se puede observar un ejemplo de este algoritmo

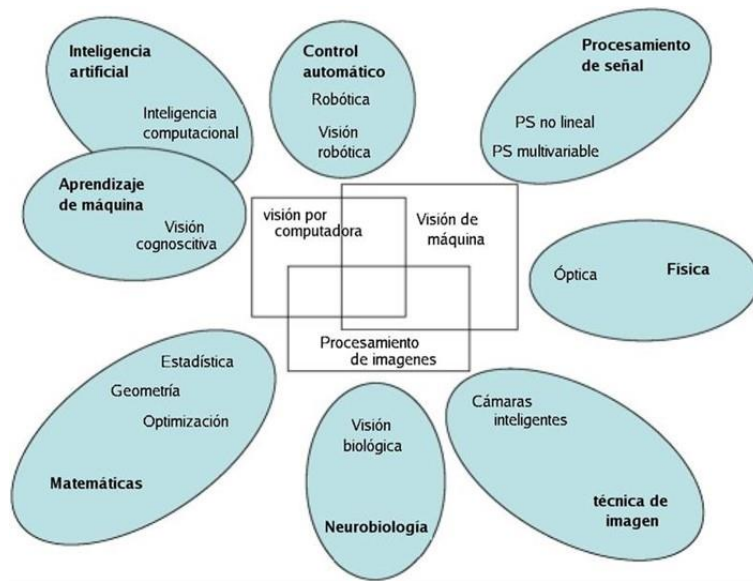


Figura 1. Panorama de las ramas de la visión por computadora.

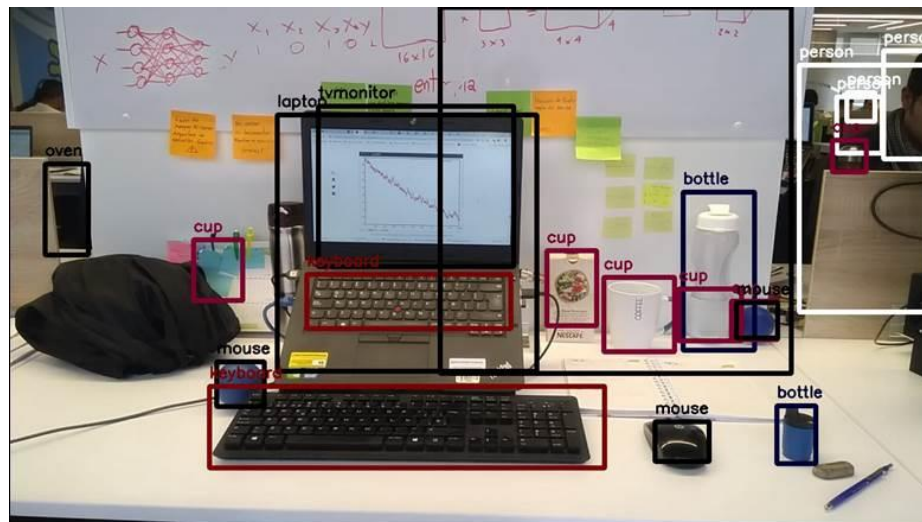


Figura 2. Algoritmo YOLO[9].

c. Redes Neuronales

Los sistemas diseñados por redes neuronales son comúnmente un modelo informático - algorítmico en el cual consiste en sistemas únicos llamados neuronas, las cuales conectadas entre sí pueden transmitir una señal eléctrica, tal y como lo hace el cerebro humano para enviar una señal a las diferentes partes del cuerpo, véase la figura 3.

“Cada neurona está conectada con otras a través de unos enlaces. En estos enlaces el valor de salida de la neurona anterior es multiplicado por un valor de peso. Estos pesos en los enlaces pueden incrementar o inhibir el estado de activación de las neuronas adyacentes” [2].

De la misma manera, a cada salida de cada una de las neuronas existe una función que restringe el valor del umbral, esta función tiene como objetivo imponer un límite a cada neurona para que respeten las fronteras de la información. A esta función se le conoce como función de activación.

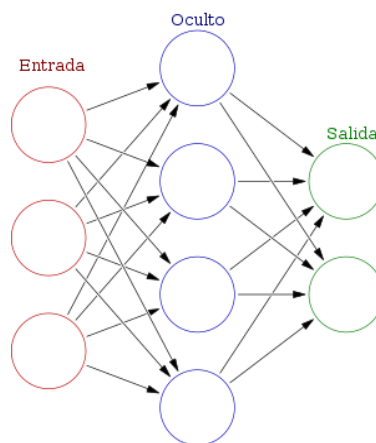


Figura 3. Diagrama de una red neuronal artificial [1].

El uso de las redes neuronales se ha diversificado en muchos campos de la ingeniería tales como los sistemas de visión, el reconocimiento de voz, la identificación de genes, entre otros en donde las técnicas de programación clásicas no pueden dar una respuesta clara o precisa.

d. TensorFlow

La librería TensorFlow es una de las más empleadas en aprendizaje máquina y aprendizaje automático, por medio de la cual es posible construir y entrenar redes neuronales que sean capaces de detectar e interpretar patrones y correlaciones muy similares al proceso que realiza una persona.

“Actualmente es utilizado tanto en la investigación como en los productos de Google frecuentemente reemplazando el rol de su predecesor de código cerrado, DistBelief. TensorFlow fue originalmente desarrollado por el equipo de Google Brain para uso interno en Google antes de ser publicado bajo la licencia de código abierto Apache” [8].

e. Motor Speech-to-Text

Un motor de conversión de discurso a texto (STT, del inglés Speech-to-Text) es una herramienta que permite procesar el habla de una persona y realizar una conversión a texto en un idioma determinado [10]. Para esto, debe de entrenarse el sistema con muestras variables de audio de discurso natural, utilizando un algoritmo de aprendizaje máquina. Para mejorar la precisión del sistema y reducir el nivel de ambigüedad del sistema de reconocimiento, se pueden definir series de comandos que deberán ser detectados en el discurso natural de una persona. En la figura 4 se muestra un diagrama de proceso de un SST

Una parte fundamental en el diseño de una SST, es que tipo de aprendizaje se debe emplear para construir los diferentes tipos de conocimiento, de estos se desprenden dos tipos:

- **Deductivo:** Estas técnicas deductivas se fundamentan en la transferencia de conocimiento humano hacia una máquina. Ejemplos de estos sistemas son los sistemas expertos
- **Inductivo:** Estos sistemas poseen un aprendizaje automático en donde los conocimientos los adquiere de forma de las bases de datos con las cuales son alimentados o de entrenamientos establecidos previamente. Ejemplo de estas técnicas son las cadenas de Márkov.

De manera práctica, no hay metodologías que solo empleen aprendizaje inductivo, por lo que actualmente existen modelos híbridos en donde el aprendizaje deductivo – inductivo, para ello estos se clasifican de la siguiente forma:

- Sistema de Entrenamiento: se determina si el sistema requiere un entrenamiento a priori al usarse el sistema
- Dependencia del interlocutor: se establece si el sistema debe ser entrenado por el usuario o este es independiente
- Continuidad: Aquí el sistema determina si este puede entender de forma continua o debe hacer pausas las instrucciones del usuario
- Robustez: Se determina si el sistema se encuentra diseñado con los suficientes filtros para evitar que señales externas al interlocutor puedan interferir, ya sea ruido ambiental o de fondo
- Tamaño del dominio: Establece si el sistema diseñado reconoce diferentes variaciones del lenguaje o expresiones coloquiales del interlocutor

f. Python

El lenguaje Python se trata de un sistema multiparadigma con el cual soporta la programación orientada a objetos, programación imperativa y la programación funcional, lo cual lo hace ser un programa dinámico y multiplataforma.

“Python es un lenguaje de programación multiparadigma. Esto significa que más que forzar a los programadores a adoptar un estilo particular de programación, permite varios estilos: programación orientada a objetos, programación imperativa y programación funcional. Otros paradigmas están soportados mediante el uso de extensiones”[15].

Una de las características mas importantes de Python es su forma dinámica en la cual los métodos y las variables pueden ser ejecutadas en cada una de las partes del

código, a esto se le llama enlace dinámico de métodos. Otra ventaja de este lenguaje es la facilidad de sus aplicaciones en interfaces programables en donde Python puede escribir módulos de forma fácil realizados en C/C++

Python es considerado como un lenguaje minimalista, donde se trata de simplificar la programación y los procesos del diseño del mismo, existen muchas analogías entre Python y los lenguajes basados en Lisp como el Scheme.

g. Tarjeta Arduino.

El Arduino Mega es una tarjeta de adquisición de datos de bajo costo basado en un microprocesador ATmega2560. Sus características técnicas son:

- Tiene 54 entradas/salidas digitales (de las cuales 15 pueden ser usadas como salidas PWM),
- 16 entradas analógicas, 4 UARTs, un cristal de 16Mhz, conexión USB, jack para alimentación DC, conector ICSP, y un botón de reset.



Figura 4. Arduino Mega.

Arduino Mega posee las siguientes especificaciones:

- **Microcontrolador:** ATmega2560
- **Voltaje Operativo:** 5V
- **Voltaje de Entrada:** 7-12V
- **Voltaje de Entrada(límites):** 6-20V
- **Pines digitales de Entrada/Salida:** 54 (de los cuales 15 proveen salida PWM)

- **Pines análogos de entrada:** 16
- **Corriente DC por cada Pin Entrada/Salida:** 40 mA
- **Corriente DC entregada en el Pin 3.3V:** 50 mA
- **Memoria Flash:** 256 KB (8KB usados por el bootloader)
- **SRAM:** 8KB
- **EEPROM:** 4KB
- **Clock Speed:** 16 MHz

1. Objetivos (general y específicos)

a. Objetivo General

Diseñar un sistema mecatrónico de asistencia en hogares para personas invidentes, basado en IoT y visión por computadora, con reconocimiento de objetos, localizador de artículos y un sistema de guía por audio.

b. Objetivo Específicos

- Diseñar el algoritmo de reconocimiento de objetos con Tensorflow.
- Entrenar el sistema por medio de aprendizaje máquina.
- Integrar el hardware con la tarjeta Arduino.
- Desarrollar el algoritmo de control del sistema.
- Implementar el motor STT (Speech-to-Text) de Google para comandos de voz.

2. Metodología

Durante la última década, los avances relacionados con las aplicaciones móviles han permitido el desarrollo de las personas dentro de la sociedad, de tal forma que el realizar diversas actividades cotidianas tales como desplazarse de un punto a otro dentro de un recinto cerrado se ha vuelto más sencillo. Sin embargo, muchas de las aplicaciones móviles que las personas utilizan hoy en día, no se desarrollan pensando en que serán utilizadas por personas que padecen algún tipo de discapacidad, como puede ser el caso de una persona invidente o con visión reducida, lo cual impide su inclusión dentro del

entorno social, afectando de manera negativa su calidad de vida. La falta de herramientas que permitan a una persona invidente el detectar fácilmente objetos que se encuentran a su alrededor dentro de un entorno cerrado, perjudica de manera negativa su desarrollo como individuos dentro de la sociedad debido al hecho de estar negando su derecho al acceso a la información y a una calidad de vida digna.

La aplicación móvil de asistencia para personas invidentes se conforma principalmente de cuatro módulos centrados en la recopilación de información, el proceso de detección de objetos, la interfaz para el sistema operativo de Android y finalmente, el algoritmo de consulta en la nube. El algoritmo para la detección de los objetos está diseñado para que al detectarse una similitud del setenta por ciento o más en un elemento dentro del campo de visión de la cámara integrada al teléfono inteligente, muestre en pantalla la clase a la que este pertenece; en caso de no detectar ninguna similitud con los elementos dentro de la base de datos de la aplicación, el elemento detectado llevara la etiqueta “Desconocido”.

La parte del algoritmo que se encarga de la detección de los objetos está diseñada a partir del modelo de visión artificial YOLO (You Only Look Once), el cual utiliza una base de datos de código libre. La base de datos que sirve como fuente principal de imágenes para detectar elemento es Google Open Images V4, esta base de datos cuenta con información referente a más de 600 diferentes clases de elementos, de los cuales únicamente se utilizaron 9 para entrenar el algoritmo de detección de la aplicación de asistencia. El subconjunto de clases utilizado corresponde a algunos de los artículos de uso cotidiano más comunes que se pueden encontrar en una casa, así como tener la característica de ser elementos que pueden ser sostenidos únicamente con una mano y puestos frente a la cámara de un dispositivo móvil. La Tabla 1 contiene el subconjunto de clases utilizado, así como una breve descripción. [9]

<i>Clases definidas para los sets de imágenes</i>		
ID	Clase	Notas
0	Persona	Diferentes tipos de personas (Edad, Género, etc....)
1	Tijeras	Herramienta de corte
2	Moneda	Diferentes tipos y tamaños
3	Taza	Diversos Materiales (Plásticos, Metal, etc....)
4	Cuchillo	Utensilio de cocina
5	Botella	Diferentes materiales (Vidrio, Plástico, etc....)
6	Calcetín	Prenda de ropa
7	Cuchara	Utensilio de cocina
8	Tenedor	Utensilio de cocina

Tabla 1 Clases extraídas de la base de datos Open Images V4

El entrenamiento del algoritmo YOLO V4 para la detección de objetos fue realizado utilizando el framework Darknet, permitiendo la clasificación e identificación de elementos en imágenes, así como su validación respecto a la base de datos, esto último permite generar métricas de desempeño y realizar tareas de detección a través de comandos en una terminal de Linux. Cabe mencionar que el entrenamiento del algoritmo incluye un proceso de retroalimentación respecto a los casos en donde exista una falla al identificar algún elemento, así como la capacidad de llegar a una capacidad de precisión más generalizada la cual permita clasificar imágenes nuevas sin ningún problema. Así mismo, Cloud storage es la herramienta que se utiliza para almacenar las imágenes que se capturen con la aplicación en la nube. El proceso de clasificación se puede observar en la Figura 5.

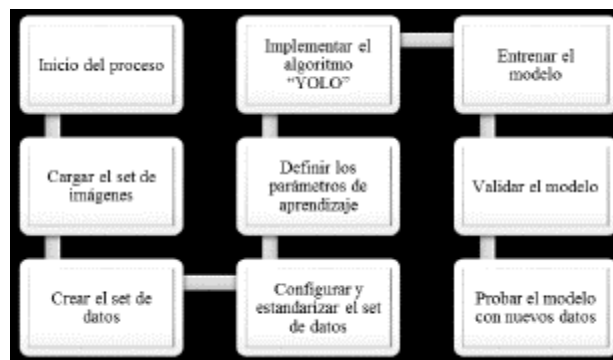


Figura51 Pasos generales del diseño del algoritmo de clasificación.

La aplicación se desarrolló pensando en ser utilizada principalmente por usuarios con dispositivos móviles que utilizaran el sistema operativo de Android, es por esta razón que se utilizaron API's del nivel 23 (equivalentes a la versión 6.0 de Android) de para el desarrollo de las funciones con las que hace interacción el algoritmo. Para la captura de imágenes se utilizó la API "CameraView", la cual consiste en un proceso que permite integrar las imágenes captadas en tiempo real por la cámara en la interfaz de la aplicación. Google ya proporciona algunos servicios para habilitar herramientas que detectan la voz del usuario en dispositivos Android, es por eso por lo que para la aplicación se utilizó la API de "SpeechRecognizer", la cual permite declarar diversos parámetros como lo son el lenguaje, la región y el tipo de consulta para que de esta forma la aplicación pueda entender de manera clara las instrucciones dadas por el usuario, al mismo tiempo la API "TextToSpeech" permite generar muestras de voz a partir de cadenas de texto utilizando los mismos parámetros. Finalmente, para que la aplicación pueda conectarse con el servicio de almacenamiento en la nube se utilizó las API's de "Cloud Firestore" y "Cloud Storage", ambas forman parte del paquete de dependencias del servicio "Firebase". El diseño de la interfaz de la aplicación se realizó utilizando la herramienta de Android Studio, esta se compone principalmente de la interfaz táctil, la cual es donde estarán ubicados los botones que permiten al usuario, ya sea realizar una consulta rápida o entrar en el modo de escaneo completo. Como se puede observar en la Figura 6, los botones están distribuidos por la pantalla del dispositivo móvil de una forma que sea sencillo recordar su ubicación para los usuarios invidentes, de la misma forma los botones poseen un tamaño considerable para asegurar que el usuario pueda presionarlo sin ningún problema.

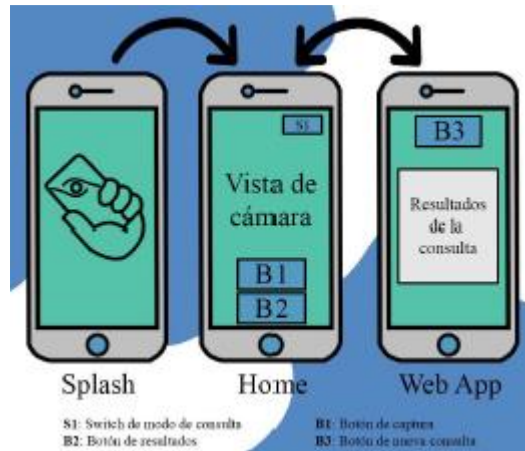


Figura 6 Modelo de interacción de la interfaz táctil.

Utilizando las API's de Text-To-Speech, se implementó la característica de brindarle al usuario mensajes que indiquen al usuario cuál es la acción que se encuentra realizando en todo momento, de esta forma el usuario puede saber cuándo debe presionar un botón al utilizar la aplicación. Al realizar una consulta utilizando la opción de actividad WEB, la aplicación toma la información de la última captura realizada y la procesa utilizando el algoritmo de reconocimiento; cuando el algoritmo identifica un objeto dentro la imagen procesada, este delimita la posición aproximada del elemento en la imagen y es marcado con una etiqueta que contiene el nombre de la clase a la que pertenece.

El sistema de consulta en la nube se desarrolló en el entorno de programación de Google Colaboratory, este se encarga de obtener la información de una imagen para después aplicar la detección de elementos utilizando el algoritmo YOLO V4. Este sistema se implementó con el objetivo de reducir el consumo de los recursos que necesita la aplicación móvil para funcionar, de esta manera se obtuvo una versión compacta la cual puede ser instalada en cualquier dispositivo con el sistema operativo Android a partir de la versión 6.0.

3. Instituciones, organismos o empresas de los sectores social, público o productivos participantes
No Aplica

4. Resultados

El diseño de esta herramienta para la asistencia de personas inválidas se desarrolló en un periodo de tiempo 14 semanas equivalentes a un semestre, esto como un requerimiento necesario para obtener la titulación de ingeniería mecatrónica. Al ser una herramienta enfocada para dispositivos móviles, su desarrollo fue completamente utilizando herramientas de código libre como es el caso del entorno de programación de Python, esto junto con el hecho de que las pruebas realizadas con el fin de validar el correcto funcionamiento de la aplicación se llevaron a cabo utilizando un dispositivo móvil de uso personas, represento un bajo costo para su desarrollo e implementación.

Un total de 15 pruebas fueron realizadas utilizando el modo consulta por clasificación, donde se colocaron frente al dispositivo móviles elementos correspondientes a las clases que pueden ser identificadas por el algoritmo de detección con un índice de confianza de sesenta por ciento en un tiempo de consulta aproximado de 6.7 segundos. En la Figura 7 se muestran las capturas hechas a diferentes elementos con el objetivo de validar el correcto funcionamiento de la aplicación al detectar e identificar diversos objetos.

Los resultados obtenidos tal y como se muestra en la Tabla 2, indican que el algoritmo fue capaz de detectar cada uno de los elementos pertenecientes a las nueve clases definidas en el desarrollo de este proyecto. Del total de registros realizados por la aplicación, se puede destacar la obtención de un falso positivo donde el algoritmo logro identificar el elemento en la imagen, pero asignándole una etiqueta errónea, de la misma forma se obtuvieron dos falsos negativos donde los elementos fueron detectados sin ningún problema, pero el algoritmo no logro identificarlos del todo otorgándoles una etiqueta de desconocidos. Los resultados restantes muestran que el algoritmo detectar elementos que no pertenecen a ninguna clase con un total de tres resultados verdaderos negativos, mientras que en los once registros restantes se logró una detección e identificación exitosa por parte de la aplicación.



Figura 7 imágenes capturadas con la aplicación de Android en modo detección

Tabla 2 Resultados obtenidos en el modo de clasificación

No.	Respuesta Esperada	Respuesta obtenida	% de confianza	Tipo de resultado	Tiempo de consulta
1	Botella	Botella	87	VP	8.42 Segundos
2	Moneda	Moneda	100	VP	9.06 Segundos
3	Taza	Taza	64	VP	7.7 Segundos
4	Persona	Persona	99	VP	8.79 Segundos
5	calcetín	Moneda	98	FP	8.8 Segundos
6	calcetín	calcetín	97	VP	7.82 Segundos
7	Desconocido	Desconocido	-	VN	8.9 Segundos
8	Tijeras	Desconocido	-	FN	9.54 Segundos
9	Tijeras	Tijeras	96	VP	8.53 Segundos
10	Tenedor	Tenedor	83	VP	6.51 Segundos
11	Botella	Desconocido	-	FN	6.23 Segundos
12	Cuchara	Cuchara	83	VP	6.7 Segundos
13	Cuchillo	Cuchillo	99	VP	6.29 Segundos
14	Desconocido	Desconocido	-	VN	6.29 Segundos
15	Desconocido	Desconocido	-	VN	6.55 Segundos

Por otro lado, las pruebas realizadas utilizando el modo de detección se llevaron a cabo poniendo la aplicación frente a diversos conjuntos de elementos físicos, así como de algunas imágenes, esto con el objetivo de verificar que el algoritmo es capaz de identificar correctamente los elementos existentes en cada caso, tal y como se muestra en la Figura 8. Los resultados muestran que dentro del tiempo de respuesta aproximado de

6.7 segundos, el porcentaje de confianza promedio fue de 91.08%, teniendo en cuenta un total de 15 consultas realizadas, dentro de las cuales, un total de 17 elementos detectados obtuvieron una respuesta falsa negativa, es decir, se detectó la presencia de un elemento en la imagen, pero no se logró identificar la clase a la que pertenece. Por otro lado, se obtuvieron un total de 57 respuestas positivas donde la detección e identificación de los elementos se logró de una forma satisfactoria.



Figura 8 Resultados obtenidos en el modo de detección

5. Productos generados

1. Tesista de ingeniería en mecatrónica.

6. Conclusiones

Al finalizar el desarrollo de este proyecto, se obtuvo un sistema capaz de funcionar en dispositivos móviles con acceso al operativo Android, el cual ofrece un algoritmo de detección que combina la tecnología de visión por computadora y el computo en la nube, lo que facilita la posibilidad de implementar actualizaciones en un futuro, a diferencia de investigaciones similares donde se hace uso de microcontroladores o microcomputadoras.

Con base en los resultados obtenidos a partir de las pruebas de funcionalidad realizadas, se puede concluir que el sistema de asistencia para personas invidentes es completamente funcional y cumple de manera satisfactoria con cada uno de los objetivos planteados en su desarrollo.

Es importante mencionar que a pesar de que la aplicación móvil funciona correctamente y es capaz de localizar e identificar elementos enfocados por el usuario, existen casos donde el algoritmo clasifica objetos de manera errónea, por lo que es necesario realizar modificaciones con el objetivo de crear un algoritmo de detección más robusto y eficiente.

El proyecto puede ser modificado en un futuro para que sean agregadas más funciones que beneficien a usuarios invidentes, de esa forma crear una herramienta cada vez más completa que ayude a personas que sufren alguna discapacidad visual a realizar otras actividades cotidianas en las que puedan presentar dificultades, las cuales pueden ser ya sea de carácter personal o laboral.

7. Mecanismos de transferencia.

No aplica.

8. Contribución e impacto del proyecto

Este proyecto representa un gran avance para la integración de las tecnologías de IOT con personas con discapacidad, en especial para las personas con algún problema de visión. El impacto principal de este proyecto es la ayuda a la asistencia de personas invidentes, las cuales son un sector vulnerable de nuestra sociedad que debido a las limitantes que estas personas pueden tener para poder realizar tareas comunes y cotidianas.

El integrar nuevas tecnologías a nuevos sectores los acerca de forma mas sencilla a la sociedad a las personas con capacidades distintas, las cuales a lo largo de la historia han demostrado tener un gran deseo de superación y de éxito. El uso del Machine Learning la cual es una herramienta de última generación y de gran interés para los investigadores, ya que actualmente esta tecnología es aplicada a diferentes sectores como la industria y la producción de bienes y servicios, por lo que la aplicación de este proyecto hacia un nuevo sector puede provocar una nueva rama de investigación hacia los lados sociales y humanos en donde la ingeniería puede auxiliar a grupos vulnerables

La elaboración del proyecto no solo es exclusiva para la carrera de ingeniera en mecatrónica, si no para la enseñanza de otras carreras de ingeniería, lo cual va a permitir favorecer la adquisición de conocimientos habilidades y valores en el proceso de

aprendizaje de distintos programas que la universidad enseña, por lo que este proyecto es de carácter multidisciplinario.

9. Impacto económico, social y/o ambiental en la región

Según datos del INEGI, “México se ubica entre los 20 países con mayor número de personas con discapacidad visual y ceguera, siendo la catarata, la retinopatía diabética, el edema macular diabético y el glaucoma las principales complicaciones responsables del deterioro visual de la población”[4].

Por lo que el impacto que tendrá este proyecto en el ámbito social es de suma importancia al ayudar a personas invidentes o con discapacidad visual al llevar una vida mucho más sencilla y digna con ayuda de la ingeniería mecatrónica y con la cual pueda resolver las cuestiones más básicas y comunes del bienestar humano. Por lo que este proyecto en vez de ser medido por un medio económico, establecemos más como un impacto social hacia un sector de la población en México que ha sido olvidado durante años en cuestiones de asistencia tanto social como tecnológica.

10. Referencias (bibliografía)

[1]Jordi Salazar and Santiago Silvestre. Internet de las cosas. Techpedia. Ceské vy-soké učení technické v Praze Fakulta elektrotechnická, 2016.

[2]Josep Prieto, Roberto Ramírez, Julián Morillo, Marc Domingo, Javier Salvador, and Helena Bolta. Tecnología y desarrollo en dispositivos móviles. Barcelona, 2017.

[3]Rodrigo Vejarano, Andrés Pitty, Gilberto Gómez, and Lineth Alain. Eye qr-aplicación identificadora de objetos para personas con discapacidad visual. Revista De Iniciación Científica, 5(2):77-82, 2019.

[4]Dulce Lourdes Loza Pacheco. Dimensión y características de la población con baja visión y ceguera en México. 2016.

[5] Carlos Aramendiz, David Escorcía, Jesús Romero, Kevyn Torres, Carlos Triana, and Silvia Moreno. Sistema basado en reconocimiento de objetos para el apoyo a personas con discapacidad visual. *Investigación y desarrollo en TIC*, 11(2):75-82, 2020.

[6] R. Phadnis, J. Mishra, and S. Bendale, “Objects Talk - Object Detection and Pattern Tracking Using TensorFlow,” *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, pp. 1216–1219, 2018.

[7] I. Joe Louis Paul, S. Sasirekha, S. Mohanavalli, C. Jayashree, P. Moohana Priya, and K. Monika, “Smart Eye for Visually Impaired-An aid to help the blind people,” *ICCIDS 2019 - 2nd International Conference on Computational Intelligence in Data Science, Proceedings, 2019*.

[8] L. G. Nick Antonopoulos, *Cloud Computing in a nutshell*, 2011, vol. 15, no. 1.

[9] M. Rouse, *Internet of Things*, available at: <https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT,2016>.

[10] D. Ghosh, S. Bhattacharya, A. Lodh, P. Mitra, D. Ghosh, and W. Mallik, “Assisted Technological Headset using Internet of Things,” *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2018*, pp. 1346–1353, 2019.

[11] S. Lenman, L. Bretzner, and B. Thuresson, “Computer vision based hand gesture inter-faces for human-computer interaction,” 04 2012.

[2] Z. Wang, D. Chen, and P. Xiao, “Design of a voice control 6dof grasping robotic arm based on ultrasonic sensor, computer vision and alexa voice assistance,” pp. 649–654, Aug 2019.

- [12] S. K. Chaturanga, K. C. Samarawickrama, H. M. L. Chandima, K. G. T. D. Chaturanga, and A. M. H. S. Abeykoon, "Hands free interface for human computer interaction," pp.359–364, 2010.
- [13] D. Lee and Y. Park, "Vision-based remote control system by motion detection and openfinger counting,"IEEE Transactions on Consumer Electronics, vol. 55, no. 4, pp. 2308–2313, 2009.
- [14] J. Hauswald, M. A. Laurenzano, Y. Zhang, C. Li, A. Rovinski, A. Khurana, R. G.Dreslinski, T. Mudge, V. Petrucci, L. Tang, and J. Mars, "Sirius: An open end-to-end voice and vision personal assistant and its implications for future warehouse scale computers,"SIGPLAN Not., vol. 50, no. 4, p. 223–238, Mar. 2015. [Online]. Available:<https://doi.org/10.1145/2775054.2694347>
- [15] I. Joe Louis Paul, S. Sasirekha, S. Mohanavalli, C. Jayashree, P. Moohana Priya, and K. Monika, "Smart eye for visually impaired-an aid to help the blind people," pp. 1–5, 2019.
- [16] R. A. Minhas and A. Javed, "X-eye: A bio-smart secure navigation framework for visually impaired people," pp. 1–4, 2018.
- [17] P. Mungkaruna, P. Piyawongwisal, K. Ropkhop, and U. Hatthasin, "The talking color identifying device for the visually impaired," pp. 1–5, 2016.

11. Anexos

11.1 Taxonomía de los Roles de Colaborador (con las actividades logradas)

Roles	Definición de los roles	Nombre de él(la) investigador(a)	Figura	Grado de contribución	Actividades logradas durante el proyecto	Tiempo promedio semanal (en horas) dedicado al proyecto
Responsabilidad de la dirección del proyecto	Coordinar la planificación y ejecución de la actividad de investigación. Organiza los roles de cada colaborador, tiene la habilidad de identificar potenciales de cada individuo para generar una sinergia de equipo colaborativo	Dr. Luis Carlos Méndez González	Director del proyecto	Principal	<ul style="list-style-type: none"> • revisión Técnica del Proyecto • revisión del Reporte técnico • supervisión del Estudiante • Planificación del cronograma de actividades 	3
Responsabilidad de supervisión	Elaborar la planificación de las actividades de la investigación (cronogramas y controles de seguimiento), describe los roles identificados por el director del proyecto y facilita el apoyo constante a todos los roles para conseguir un trabajo integral, coherente y que llegue a buen término	Dr. Luis Alberto Rodríguez Picón	Supervisor	Principal	<ul style="list-style-type: none"> • revisión Técnica del Proyecto • revisión del Reporte técnico • supervisión del Estudiante • Planificación del cronograma de actividades 	3
Realización y redacción de la propuesta	Preparación, creación y redacción de la propuesta de investigación, específicamente la redacción, revisión de coherencia del texto, presentación de los datos y la normatividad aplicable para garantizar el cumplimiento de los requisitos.	Dr. Iván Juan Carlos Pérez Olgún	Redactor de la propuesta	Principal	<ul style="list-style-type: none"> • redacción de la propuesta del proyecto • Revisión técnica del proyectos 	2

Desarrollo o diseño de la metodología	Contribuir con el diseño de la metodología, modelos a implementar y el sustento teórico, empírico y científico para la aplicabilidad de los instrumentos en la ejecución del proyecto.	Dr. David Luviano Cruz	Desarrollo de la metodología	Principal	<ul style="list-style-type: none"> Desarrollador de la metodología del proyecto 	3
Desarrollo o diseño de la metodología	Contribuir con el diseño de la metodología, modelos a implementar y el sustento teórico, empírico y científico para la aplicabilidad de los instrumentos en la ejecución del proyecto.	Dr. Luis Asuncion Pérez Dominguez	Desarrollo de la metodología	Principal	<ul style="list-style-type: none"> Desarrollador de la metodología del proyecto 	3
Recopilación/Recolección de datos	Ejecuta las estrategias propuestas en acciones encaminadas a obtener la información, haciendo la recopilación de datos y la inclusión de la evidencia en el proceso.	Mtro. Abel Eduardo Quezada Carreón	Recopilador de datos	Apoyo	<ul style="list-style-type: none"> Recolección de datos para validar la metodología 	2
Recopilación/Recolección de datos	Aplicar métodos estadísticos, matemáticos, computacionales, teóricos u otras técnicas formales para analizar o sintetizar los datos del estudio. Verifica los resultados preliminares de cada etapa del análisis, los experimentos implementados y otros productos comprometidos en el proyecto.	Mtra. Luz Angelica García Villalba	Recopilación de datos	Apoyo	<ul style="list-style-type: none"> validación de la metodología propuesta 	2
Elaboración del análisis formal de la investigación	Aplicar métodos estadísticos, matemáticos, computacionales, teóricos u otras técnicas formales para analizar o sintetizar los datos del estudio. Verifica los resultados preliminares de cada etapa del análisis, los experimentos implementados y otros productos comprometidos en el proyecto.	Mtra. Fabiola Lom Monárrez	Análisis de datos	Apoyo	<ul style="list-style-type: none"> Derivar conclusiones de los datos recabados, así como su interpretación matemática 	2

Preparación, creación y/o presentación de los productos o entregables	Preparar la redacción del reporté técnico de avance parcial y el reporte técnico final. Se hace la revisión crítica, la recopilación de las observaciones y comentarios del grupo de investigación. Y finalmente se procede a la edición del documento a entregar.	Mtro. Héctor Loya Caraveo.	Editor de reportes técnicos	Apoyo	<ul style="list-style-type: none"> • Redacción de los reportes técnicos del proyecto • Revisión de la tesis del estudiante 	2
--	--	----------------------------	-----------------------------	-------	--	---

11.1.1 Estudiantes participantes en el proyecto

Nombre de estudiante(s)	Matrícula	Tiempo promedio semanal (en horas) dedicado al proyecto	Actividades logradas en la ejecución del proyecto
ALAN IVÁN HERNÁNDEZ HOLGUÍN	159799	4	<ul style="list-style-type: none"> • Recopilación de los datos necesarios para probar la metodología • Diseño del software propuesto en la metodología • Validación de la metodología

Acta de Examen de Grado del Estudiante



Universidad Autónoma de Ciudad Juárez

Instituto de Ingeniería y Tecnología – Ciudad Universitaria
Departamento de Industrial y Manufactura


ACTA DE EXAMEN DE PROYECTO DE TITULACIÓN


Ciudad Juárez, Chihuahua a 28 del mes de mayo del 2021, siendo las 12:00 horas, se reúnen a través de la plataforma virtual de Microsoft Teams, los miembros del Jurado:


Presidente: Dr. Luis Alberto Rodríguez Picón
Sinodal (1): Dr. Iván Juan Carlos Pérez Olguín
Sinodal (2): Dra. Adriana Salinas Ávila
Lector de tesis: MC. Luz Angelica García Villalba

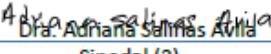
Se procede a efectuar la evaluación del proyecto de titulación que presenta el C. **ALAN IVÁN HERNÁNDEZ HOLGUÍN** con matrícula **159779**, alumno del programa de Ingeniería en Mecatrónica. Tomando en cuenta los miembros del Jurado el contenido del proyecto cuyo tema es: **“SISTEMA AUXILIAR EN HOGARES PARA PERSONAS INVIDENTES BASADO EN IOT”** dirigido por Dr. Luis Carlos Méndez González enfocando a personal y la réplica del mismo dictamina que es **APROBADO**. El presidente del Jurado hace saber al sustentante el resultado obtenido.


Se da por terminado el acto a las 12:50 horas, y una vez escrita, leída y aprobada la firma para dar constancia, las personas que en el acto intervienen.

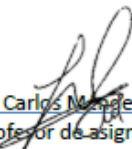

Dr. Luis Alberto Rodríguez Picón
Presidente


Dr. Luis Carlos Méndez González
Director del proyecto


Dr. Iván Juan Carlos Pérez Olguín
Sinodal (1)


Dra. Adriana Salinas Ávila
Sinodal (2)


MC. Luz Angelica García Villalba
Lector de Tesis


Dr. Luis Carlos Méndez González
Profesor de asignatura

DEPARTAMENTO DE INGENIERÍA INDUSTRIAL Y MANUFACTURA



Aplicación Móvil para Asistencia de Personas Invidentes

Santiago Elías Rodríguez Martínez¹, Dr. Luis Carlos Méndez González²,
Dr. Luis Alberto Rodríguez Picón³ y Dr. Iván Juan Carlos Pérez Olguín⁴

Resumen—Se desarrolló una aplicación para dispositivos móviles la cual combina el cómputo en la nube con un sistema de visión artificial que, mediante una interfaz diseñada para reaccionar a comandos de voz, es capaz de identificar objetos dentro de un entorno cerrado. El producto final es en una aplicación para el sistema operativo Android que permite a usuarios invidentes detectar elementos mediante la cámara un dispositivo móvil, las imágenes capturadas se comparan con una base de datos, de manera que la interfaz delimita la posición del objeto detectado además de generar una etiqueta que indica la clase a la que este pertenece. Los resultados obtenidos permiten concluir que, el sistema es capaz de realizar con éxito consultas por medio de comandos de voz para identificar elementos a partir de la información almacenada en la base de datos con un porcentaje de confianza satisfactorio y un tiempo de consulta apropiado.

Palabras clave—Tecnologías asistenciales, Aplicación móvil, Visión artificial, Discapacidad Visual, Localización en interiores.

Introducción

El concepto de Internet de las cosas IoT (Por sus siglas en inglés Internet of Things) es algo que ha estado presente en muchos de los aspectos de la vida cotidiana de las personas por algo más de tres décadas, prácticamente desde el inicio mismo de la Internet, sin embargo, en la actualidad ha tomado una mayor relevancia con el mercado creciente de las aplicaciones para dispositivos inteligentes. La parte del “Internet” es lo que permite a múltiples dispositivos electrónicos estar interconectados entre ellos y de esta forma tener la posibilidad de acceder a una gran cantidad de información, mientras que el aspecto de “las cosas” hace referencia a todo aquel elemento físico con el que una persona tiene contacto en su vida diaria, ya sea su computadora, su automóvil, o incluso su propio teléfono móvil. En general, todos estos elementos comparten la característica de contar con sensores o circuitos integrados que permiten el manejo de datos, ya sea con otros dispositivos cercanos o mediante una conexión a Internet. (Salazar y Silvestre, 2016)

En años recientes, uno de los mayores retos para la tecnología, consiste en poder superar las brechas que son ocasionadas a causa de una discapacidad, ya sea social, laboral o incluso emocional. En el caso de una discapacidad visual, el principal problema al que se enfrentan las personas es tener que depender de forma permanente de una herramienta que les permita desarrollar sus actividades diarias con normalidad, como lo suelen ser los bastones, perros lazarillos y sistemas de señalamiento para personas invidentes como por ejemplo el Braille.

De la misma forma, el mercado actual de teléfonos inteligentes, así como el de las aplicaciones para cada sistema operativo, no suelen estar enfocadas a personas que padezcan algún problema visual, por lo que las herramientas que se pueden encontrar suelen tener una gran cantidad de limitaciones en cuanto a las funciones que ofrecen, además de que comúnmente suelen ser servicios de paga o de acceso limitado especialmente a la región en la que se desarrolló. (Prieto et al., 2017)

La visión artificial tiene como objetivo el reproducir el sentido de la vista de las personas e implementarlo en una máquina de tal forma que, a partir de la información obtenida de una imagen, el algoritmo del sistema decida qué acción se deberá de realizar a continuación. Por la parte del hardware de un sistema de visión se suelen utilizar sensores que sean capaces de detectar, procesar y analizar imágenes en tiempo real como lo es en el caso de una cámara digital; mientras que para el software del sistema se necesita de un algoritmo que sea capaz de analizar y procesar la información proveniente de una imagen, para de esta forma poder crear una descripción detallada del entorno. (Vejarano et al., 2019)

¹ Santiago Elías Rodríguez Martínez es estudiante de la carrera de ingeniería en Mecatrónica de la Universidad Autónoma de Ciudad Juárez. al159696@alumnos.uacj.mx

² Dr. Luis Carlos Méndez González es profesor investigador del departamento de ingeniería industrial y manufactura en la Universidad Autónoma de Ciudad Juárez. luis.mendez@uacj.mx

³ Dr. Luis Alberto Rodríguez Picón es profesor investigador del departamento de ingeniería industrial y manufactura en la Universidad Autónoma de Ciudad Juárez. luis.picon@uacj.mx

⁴ Dr. Iván Juan Carlos Pérez Olguín es profesor investigador del departamento de ingeniería industrial y manufactura en la Universidad Autónoma de Ciudad Juárez. ivan.perez@uacj.mx

Descripción del Método

Durante la última década, los avances relacionados con las aplicaciones móviles han permitido el desarrollo de las personas dentro de la sociedad, de tal forma que el realizar diversas actividades cotidianas tales como desplazarse de un punto a otro dentro de un recinto cerrado se ha vuelto más sencillo. Sin embargo, muchas de las aplicaciones móviles que las personas utilizan hoy en día, no se desarrollan pensando en que serán utilizadas por personas que padecen algún tipo de discapacidad, como puede ser el caso de una persona invidente o con visión reducida, lo cual impide su inclusión dentro del entorno social, afectando de manera negativa su calidad de vida. La falta de herramientas que permitan a una persona invidente el detectar fácilmente objetos que se encuentran a su alrededor dentro de un entorno cerrado, perjudica de manera negativa su desarrollo como individuos dentro de la sociedad debido al hecho de estar negando su derecho al acceso a la información y a una calidad de vida digna. (Loza, 2016)

La aplicación móvil de asistencia para personas invidentes se conforma principalmente de cuatro módulos centrados en la recopilación de información, el proceso de detección de objetos, la interfaz para el sistema operativo de Android y finalmente, el algoritmo de consulta en la nube. El algoritmo para la detección de los objetos está diseñado para que al detectarse una similitud del setenta por ciento o más en un elemento dentro del campo de visión de la cámara integrada al teléfono inteligente, muestre en pantalla la clase a la que este pertenece; en caso de no detectar ninguna similitud con los elementos dentro de la base de datos de la aplicación, el elemento detectado llevara la etiqueta “Desconocido”.

La parte del algoritmo que se encarga de la detección de los objetos está diseñada a partir del modelo de visión artificial YOLO (You Only Look Once), el cual utiliza una base de datos de código libre. La base de datos que sirve como fuente principal de imágenes para detectar elemento es Google Open Images V4, esta base de datos cuenta con información referente a más de 600 diferentes clases de elementos, de los cuales únicamente se utilizaron 9 para entrenar el algoritmo de detección de la aplicación de asistencia. El subconjunto de clases utilizado corresponde a algunos de los artículos de uso cotidiano más comunes que se pueden encontrar en una casa, así como tener la característica de ser elementos que pueden ser sostenidos únicamente con una mano y puestos frente a la cámara de un dispositivo móvil. La Tabla 1 contiene el subconjunto de clases utilizado, así como una breve descripción. (Aramendiz et al., 2020)

<i>Clases definidas para los sets de imágenes</i>		
ID	Clase	Notas
0	Persona	Diferentes tipos de personas (Edad, Género, etc....)
1	Tijeras	Herramienta de corte
2	Moneda	Diferentes tipos y tamaños
3	Taza	Diversos Materiales (Plásticos, Metal, etc....)
4	Cuchillo	Utensilio de cocina
5	Botella	Diferentes materiales (Vidrio, Plástico, etc....)
6	Calcetín	Prenda de ropa
7	Cuchara	Utensilio de cocina
8	Tenedor	Utensilio de cocina

Tabla 1 Clases extraídas de la base de datos Open Images V4

El entrenamiento del algoritmo YOLO V4 para la detección de objetos fue realizado utilizando el framework Darknet, permitiendo la clasificación e identificación de elementos en imágenes, así como su validación respecto a la base de datos, esto último permite generar métricas de desempeño y realizar tareas de detección a través de comandos en una terminal de Linux. Cabe mencionar que el entrenamiento del algoritmo incluye un proceso de retroalimentación respecto a los casos en donde exista una falla al identificar algún elemento, así como la capacidad de llegar a una capacidad de precisión más generalizada la cual permita clasificar imágenes nuevas sin ningún problema. Así mismo, Cloud storage es la herramienta que se utiliza para almacenar las imágenes que se capturen con la aplicación en la nube. El proceso de clasificación se puede observar en la Figura 1.

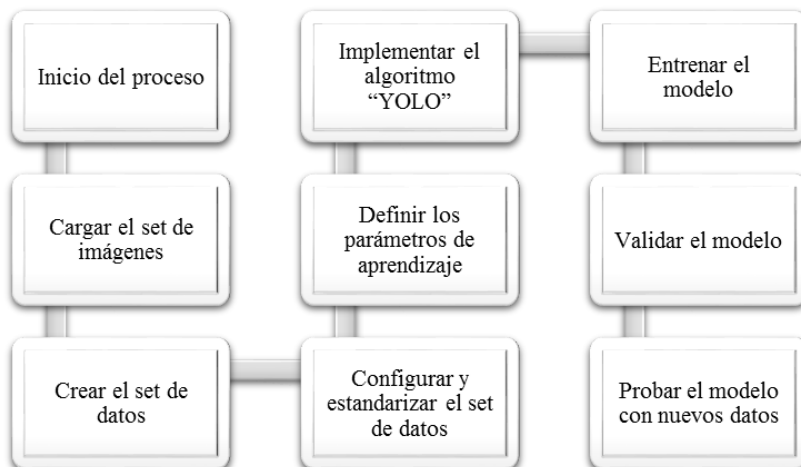


Figura 1 Pasos generales del diseño del algoritmo de clasificación

La aplicación se desarrolló pensando en ser utilizada principalmente por usuarios con dispositivos móviles que utilizaran el sistema operativo de Android, es por esta razón que se utilizaron API's del nivel 23 (equivalentes a la versión 6.0 de Android) de para el desarrollo de las funciones con las que hace interacción el algoritmo. Para la captura de imágenes se utilizó la API "CamaraView", la cual consiste en un proceso que permite integrar las imágenes captadas en tiempo real por la cámara en la interfaz de la aplicación. Google ya proporciona algunos servicios para habilitar herramientas que detectan la voz del usuario en dispositivos Android, es por eso por lo que para la aplicación se utilizó la API de "SpeechRecognizer", la cual permite declarar diversos parámetros como lo son el lenguaje, la región y el tipo de consulta para que de esta forma la aplicación pueda entender de manera clara las instrucciones dadas por el usuario, al mismo tiempo la API "TextToSpeech" permite generar muestras de voz a partir de cadenas de texto utilizando los mismos parámetros. Finalmente, para que la aplicación pueda conectarse con el servicio de almacenamiento en la nube se utilizó las API's de "Cloud Firestore" y "Cloud Storage", ambas forman parte del paquete de dependencias del servicio "Firebase". El diseño de la interfaz de la aplicación se realizó utilizando la herramienta de Android Studio, esta se compone principalmente de la interfaz táctil, la cual es donde estarán ubicados los botones que permiten al usuario, ya sea realizar una consulta rápida o entrar en el modo de escaneo completo. Como se puede observar en la Figura 2, los botones están distribuidos por la pantalla del dispositivo móvil de una forma que sea sencillo recordar su ubicación para los usuarios invidentes, de la misma forma los botones poseen un tamaño considerable para asegurar que el usuario pueda presionarlo sin ningún problema.

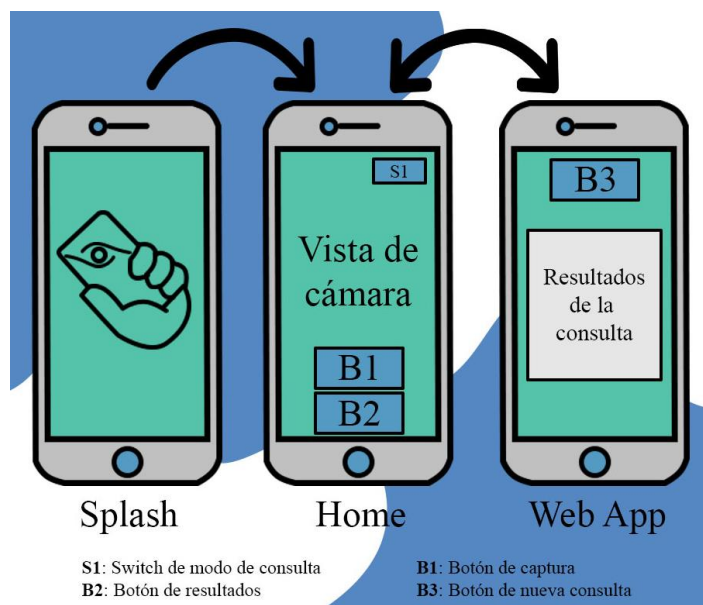


Figura 2 Modelo de interacción de la interfaz táctil

Utilizando las API's de Text-To-Speech, se implementó la característica de brindarle al usuario mensajes que indiquen al usuario cuál es la acción que se encuentra realizando en todo momento, de esta forma el usuario puede saber cuándo debe presionar un botón al utilizar la aplicación. Al realizar una consulta utilizando la opción de actividad WEB, la aplicación toma la información de la última captura realizada y la procesa utilizando el algoritmo de reconocimiento; cuando el algoritmo identifica un objeto dentro la imagen procesada, este delimita la posición aproximada del elemento en la imagen y es marcado con una etiqueta que contiene el nombre de la clase a la que pertenece.

El sistema de consulta en la nube se desarrolló en el entorno de programación de Google Colaboratory, este se encarga de obtener la información de una imagen para después aplicar la detección de elementos utilizando el algoritmo YOLO V4. Este sistema se implementó con el objetivo de reducir el consumo de los recursos que necesita la aplicación móvil para funcionar, de esta manera se obtuvo una versión compacta la cual puede ser instalada en cualquier dispositivo con el sistema operativo Android a partir de la versión 6.0.

Resultados

El diseño de esta herramienta para la asistencia de personas inválidas se desarrolló en un periodo de tiempo 14 semanas equivalentes a un semestre, esto como un requerimiento necesario para obtener la titulación de ingeniería mecatrónica. Al ser una herramienta enfocada para dispositivos móviles, su desarrollo fue completamente utilizando herramientas de código libre como es el caso del entorno de programación de Python, esto junto con el hecho de que las pruebas realizadas con el fin de validar el correcto funcionamiento de la aplicación se llevaron a cabo utilizando un dispositivo móvil de uso personas, represento un bajo costo para su desarrollo e implementación.

Un total de 15 pruebas fueron realizadas utilizando el modo consulta por clasificación, donde se colocaron frente al dispositivo móviles elementos correspondientes a las clases que pueden ser identificadas por el algoritmo de detección con un índice de confianza de sesenta por ciento en un tiempo de consulta aproximado de 6.7 segundos. En la Figura 3 se muestran las capturas hechas a diferentes elementos con el objetivo de validar el correcto funcionamiento de la aplicación al detectar e identificar diversos objetos.

Los resultados obtenidos tal y como se muestra en la Tabla 2, indican que el algoritmo fue capaz de detectar cada uno de los elementos pertenecientes a las nueve clases definidas en el desarrollo de este proyecto. Del total de registros realizados por la aplicación, se puede destacar la obtención de un falso positivo donde el algoritmo logro identificar el elemento en la imagen, pero asignándole una etiqueta errónea, de la misma forma se obtuvieron dos falsos negativos donde los elementos fueron detectados sin ningún problema, pero el algoritmo no logro identificarlos del todo otorgándoles una etiqueta de desconocidos. Los resultados restantes muestran que el algoritmo detectar elementos que no pertenecen a ninguna clase con un total de tres resultados verdaderos negativos, mientras que en los once registros restantes se logró una detección e identificación exitosa por parte de la aplicación.



Figura 3 imágenes capturadas con la aplicación de Android en modo detección

Resultados de modo clasificación					
No.	Respuesta Esperada	Respuesta obtenida	% de confianza	Tipo de resultado	Tiempo de consulta
1	Botella	Botella	87	VP	8.42 Segundos
2	Moneda	Moneda	100	VP	9.06 Segundos
3	Taza	Taza	64	VP	7.7 Segundos
4	Persona	Persona	99	VP	8.79 Segundos
5	calcetín	Moneda	98	FP	8.8 Segundos
6	calcetín	calcetín	97	VP	7.82 Segundos
7	Desconocido	Desconocido	-	VN	8.9 Segundos
8	Tijeras	Desconocido	-	FN	9.54 Segundos
9	Tijeras	Tijeras	96	VP	8.53 Segundos
10	Tenedor	Tenedor	83	VP	6.51 Segundos
11	Botella	Desconocido	-	FN	6.23 Segundos
12	Cuchara	Cuchara	83	VP	6.7 Segundos
13	Cuchillo	Cuchillo	99	VP	6.29 Segundos
14	Desconocido	Desconocido	-	VN	6.29 Segundos
15	Desconocido	Desconocido	-	VN	6.55 Segundos

Tabla 2 Resultados obtenidos en el modo de clasificación

Por otro lado, las pruebas realizadas utilizando el modo de detección se llevaron a cabo poniendo la aplicación frente a diversos conjuntos de elementos físicos, así como de algunas imágenes, esto con el objetivo de verificar que el algoritmo es capaz de identificar correctamente los elementos existentes en cada caso, tal y como se muestra en la Figura 4. Los resultados muestran que dentro del tiempo de respuesta aproximado de 6.7 segundos, el porcentaje de confianza promedio fue de 91.08%, teniendo en cuenta un total de 15 consultas realizadas, dentro de las cuales, un total de 17 elementos detectados obtuvieron una respuesta falsa negativa, es decir, se detectó la presencia de un elemento en la imagen, pero no se logró identificar la clase a la que pertenece. Por otro lado, se obtuvieron un total de 57 respuestas positivas donde la detección e identificación de los elementos se logró de una forma satisfactoria.

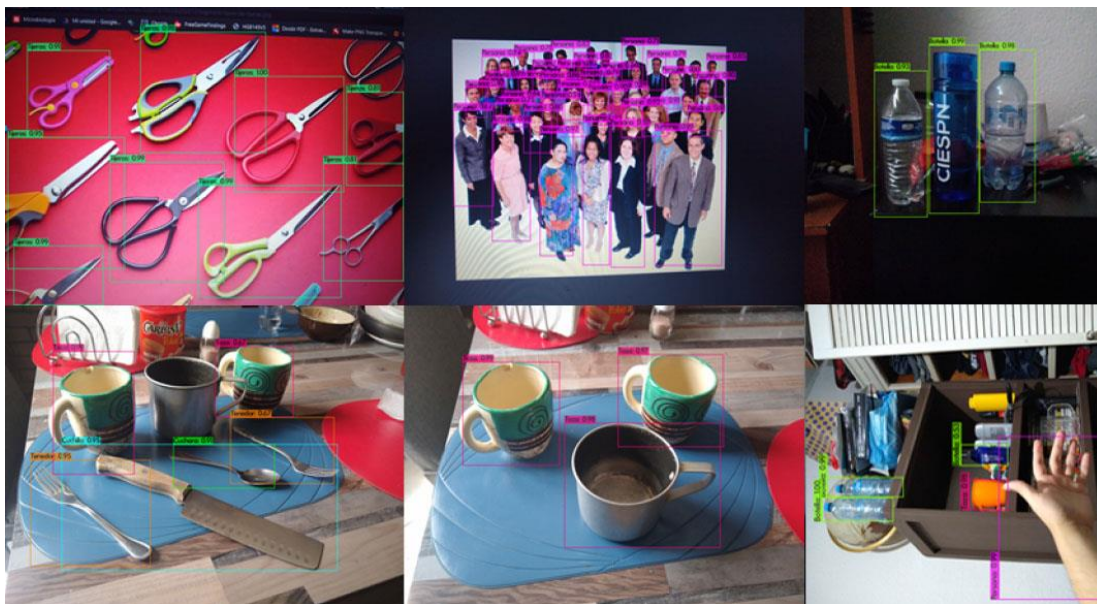


Figura 4 Resultados obtenidos en el modo de detección

Conclusiones

Al finalizar el desarrollo de este proyecto, se obtuvo un sistema capaz de funcionar en dispositivos móviles con acceso al operativo Android, el cual ofrece un algoritmo de detección que combina la tecnología de visión por computadora y el computo en la nube, lo que facilita la posibilidad de implementar actualizaciones en un futuro, a diferencia de investigaciones similares donde se hace uso de microcontroladores o microcomputadoras.

Con base en los resultados obtenidos a partir de las pruebas de funcionalidad realizadas, se puede concluir que el sistema de asistencia para personas invidentes es completamente funcional y cumple de manera satisfactoria con cada uno de los objetivos planteados en su desarrollo.

Es importante mencionar que a pesar de que la aplicación móvil funciona correctamente y es capaz de localizar e identificar elementos enfocados por el usuario, existen casos donde el algoritmo clasifica objetos de manera errónea, por lo que es necesario realizar modificaciones con el objetivo de crear un algoritmo de detección más robusto y eficiente.

El proyecto puede ser modificado en un futuro para que sean agregadas más funciones que beneficien a usuarios invidentes, de esa forma crear una herramienta cada vez más completa que ayude a personas que sufren alguna discapacidad visual a realizar otras actividades cotidianas en las que puedan presentar dificultades, las cuales pueden ser ya sea de carácter personal o laboral.

Referencias

Jordi Salazar and Santiago Silvestre. Internet de las cosas. Techpedia. České vy-soké učené technické v Praze Fakulta elektrotechnická, 2016.

Josep Prieto, Roberto Ramírez, Julián Morillo, Marc Domingo, Javier Salvador, and Helena Bolta. Tecnología y desarrollo en dispositivos móviles. Barcelona, 2017.

Rodrigo Vejarano, Andrés Pitty, Gilberto Gómez, and Lineth Alain. Eye qr-aplicación identificadora de objetos para personas con discapacidad visual. Revista De Iniciación Científica, 5(2):77{82, 2019.

Dulce Lourdes Loza Pacheco. Dimensión y características de la población con baja visión y ceguera en México. 2016.

Carlos Aramendiz, David Escorcía, Jesús Romero, Kevyn Torres, Carlos Triana, and Silvia Moreno. Sistema basado en reconocimiento de objetos para el apoyo a personas con discapacidad visual. Investigación y desarrollo en TIC, 11(2):75{82, 2020.