

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.

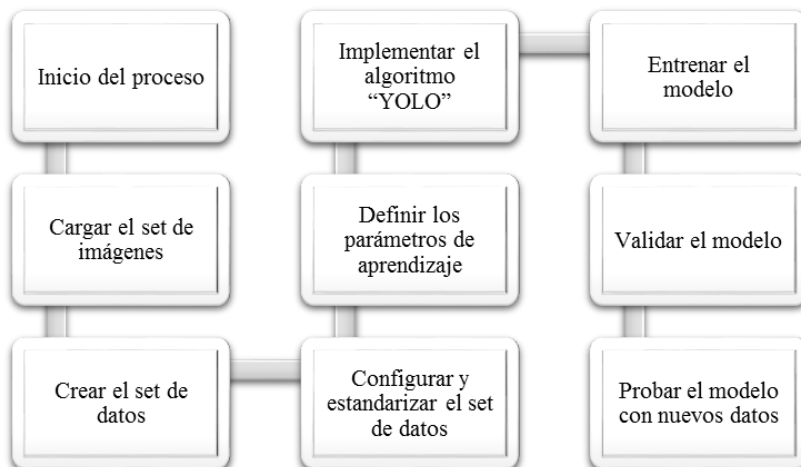


Figura 1 Pasos generales del diseño del algoritmo de clasificación

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.

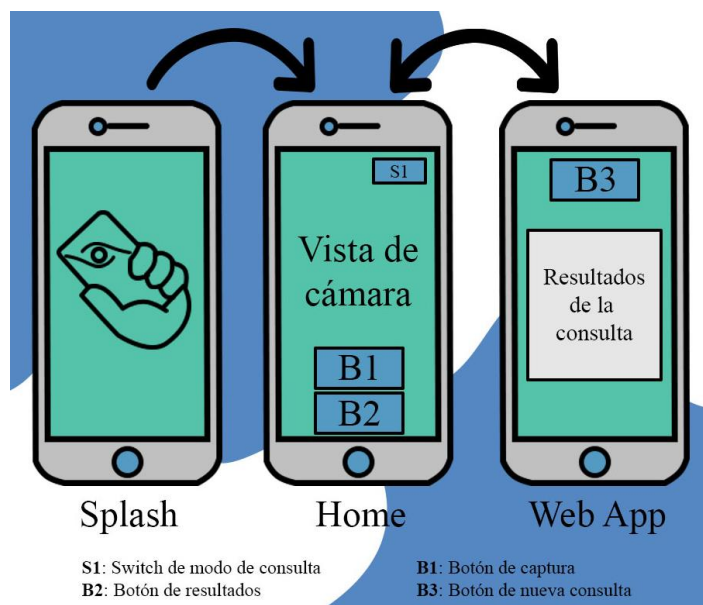


Figura 2 Modelo de interacción de la interfaz táctil

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.

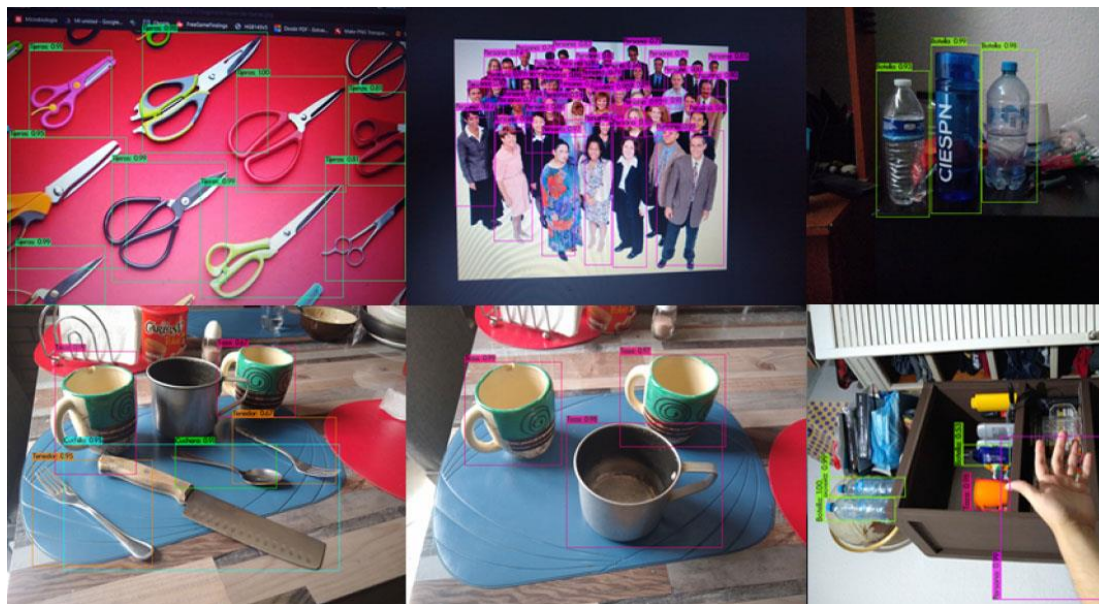


Figura 4 Resultados obtenidos en el modo de detección

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.