

Aplicación para la Asistencia Inteligente a Personas con Discapacidad Visuales

Application for the Intelligent Assistance to Visual Impaired People

Maikel Salas Zaldívar¹

*Autor de correspondencia: maikelsz8143@nauta.cu

Resumen

La vista es uno de los sentidos de mayor importancia para el ser humano por lo que las personas con discapacidad visual afrontan dificultades para su desenvolvimiento día a día. Sin embargo, muchas de estas personas poseen dispositivos móviles, que, si bien su forma de interacción no está explícitamente diseñada para estas ser utilizado por estas personas, logran interactuar con ellos gracias a tecnologías de accesibilidad como lectores de pantalla y sintetizadores de voz. Estos dispositivos en su mayoría poseen una cámara digital que permite capturar imagen y video, así como otros sensores. Haciendo uso de estas capacidades y mediante la aplicación de técnicas de procesamiento de imagen, reconocimiento de patrones e inteligencia artificial, se ha desarrollado una aplicación, que mejorando otras anteriormente creadas y funcionando en móviles y tabletas de la plataforma Android que les permite a las personas con discapacidades visuales reconocer billetes, tarjetas bancarias, objetos, reconocer el entorno y navegarlo, así como colores y niveles de iluminación. Toda la interacción con la aplicación se realiza por vía táctil, y la realimentación es mediante voz sintetizada, sonidos o reacciones táctiles. De esta manera estas personas ganan en independencia y mejora su inserción social.

Palabras clave: aplicación, asistencia, discapacitado, inteligencia artificial, dispositivo móvil.

Abstract

Sight is one of the most important senses for humans, which is why people with visual disabilities face difficulties in their daily functioning. However, many of these people have mobile devices, which, although their form of interaction is not explicitly designed for use by these people, they manage to interact with them thanks to accessibility technologies such as screen readers and voice synthesizers. These devices mostly have a digital camera that allows them to capture images and videos, as well as other sensors. Using these capabilities and through the application of image processing techniques, pattern recognition and artificial intelligence, an application has been developed, which, improving others previously created and running on mobile phones and tablets on the Android platform, allows people with visual disabilities to recognize bills, bank cards, objects, codes, recognize the environment and navigate it, as well as colors and lighting levels. All interaction with the application is done via touch, and feedback is through synthesized voice, sounds or tactile reactions. In this way, these people gain independence and improve their social integration.

Keywords: application, assistance, visual impairment, artificial intelligence, mobile.

1. Introducción

La vista es un sentido de mucha importancia para el ser humano que nos permite desenvolvernó con mayor facilidad en las actividades diarias. Según estimaciones de la Organización Mundial de la Salud (OMS), en el mundo había aproximadamente 285 millones de personas (más del 4% de la población mundial) con alguna discapacidad visual, de las cuales 39 millones eran ciegas [1, 2]. En el caso de Cuba, hay más de 30 mil personas ciegas o con discapacidad visual asociadas a la ANCI (Asociación Nacional del Ciego).

El creciente número de teléfonos móviles los han convertido en el principal medio de comunicación y fuente de información. Sin embargo, los usuarios con discapacidad visual no pueden utilizar todas las aplicaciones que se desarrollan. A diferencia de los usuarios videntes enfrentan una obvia dificultad para interactuar con teléfonos móviles de pantalla táctil. La inaccesibilidad de los teléfonos móviles se debe principalmente a sus técnicas de interacción, que requieren que el usuario ubique visualmente los objetos en la pantalla [3].

Para garantizar que las aplicaciones de software y los sitios web sean accesibles para las personas con discapacidad visual, existen una serie de pautas de accesibilidad que los desarrolladores deben seguir. Por ejemplo, las Pautas de accesibilidad al contenido web [4] brindan recomendaciones detalladas para hacer que los sitios web sean accesibles para los usuarios con discapacidades visuales. Estas pautas cubren desde el contraste de color hasta la navegación del teclado y se tienen en cuenta posibles problemas como pantallas táctiles, pequeñas, diferentes formas de interacción, así como comportamiento en diversos entornos.

Las personas discapacitadas visuales deben afrontar diariamente diferentes situaciones que generan tensión física y psicológica condicionando su calidad de vida [5]. La orientación y comprensión del entorno más allá del alcance de su baja visión, del bastón asistente o de la capacidad auditiva es una dificultad para los invidentes. Otra dificultad está en los obstáculos que encuentran durante la marcha, con los que pueden tropezar e incluso provocarles accidentes. Otras dificultades son el conocer el color de un objeto o el nivel de iluminación local. Otra es reconocer y diferenciar correctamente la denominación de los billetes o los tipos de tarjetas magnéticas o similares.

Los usuarios discapacitados visuales han tenido que crear métodos propios para utilizar teléfonos móviles o encontrar alternativas. Este problema puede resolverse con tecnologías de asistencia que permiten a estos usuarios acceder a los teléfonos móviles más fácilmente [6, 7]. Hay muchos tipos de tecnología de asistencia disponibles para ayudar a las personas con discapacidad visual a interactuar con las computadoras. Por ejemplo, los lectores de pantalla, que son programas de software que leen en voz alta el texto de la pantalla, pueden ayudar a los usuarios con discapacidades visuales a navegar a través de sitios web y aplicaciones de software. También hay pantallas braille, que utilizan una serie de pines para mostrar texto en braille, así como software de reconocimiento de voz y otras herramientas. Así los usuarios ciegos pueden conocer qué ocurre en la pantalla y con qué interactúan, y de esta manera utilizar muchas más aplicaciones de software.

Existen algunas aplicaciones especializadas de asistencia a las personas con discapacidad visual que hacen uso de funciones avanzadas de procesamiento de imagen, reconocimiento de patrones e inteligencia artificial aplicadas a la asistencia [8]. Algunas están disponibles en tiendas de software como Google Play para Android o App Store para iOS, otras en la web propia del desarrollador, otras son solo resultado de desarrollos experimentales. Algunos ejemplos son: Currency Identifier, Money Reader, ViaOptaDaily, Blind-DroidWallet, TapTapSee, NavEye, NavCog, Supersense, Smart Backpack, Sightless Helper, Google Lookout, Microsoft SeeingAI. Algunas de estas aplicaciones van destinadas a reconocer papel moneda, navegar el entorno, reconocer textos, conectar con personas asistentes voluntarios, entre otras. Usualmente implementan una única función, por ejemplo, reconocer billetes y pocas tienen algunas adicionales. Estas funciones disjuntas pueden obligar a los usuarios a tener que disponer de dos o más aplicaciones si quieren

cubrir la mayor cantidad de necesidades de asistencia. Implica para ellos tener que pasar de unas aplicaciones a otras, las que pueden tener diferentes maneras de funcionar o interactuar.

Google Lookout y Microsoft SeeingAI son las más completas y avanzadas, aunque en nuestra opinión, también adolecen de varias limitantes.

- SeeingAI solo funciona en el sistema operativo iOS. En Cuba, este tiene escasa presencia.
- Poseen interfaces visuales, innecesarias o inconvenientes para sus usuarios objetivo.
- En el caso de reconocer billetes, no incluyen ni monedas ni tarjetas cubanas.
- Algunas requieren conectividad a servicios de terceros, con posibles intermitencias, y consideraciones de costes (tráfico de datos) y de privacidad.

Esto llevó inicialmente al desarrollo de una aplicación que, ejecutándose también en dispositivos móviles, cubriera algunas necesidades de las personas discapacitadas ajustadas a nuestro contexto nacional [9, 10]. La evolución y adopción de más y mejores técnicas de análisis e inteligencia sobre imagen dio paso a concebir, investigar y desarrollar una nueva y más potente aplicación, que cubriera la mayor cantidad de posibles necesidades de asistencia de las personas con discapacidades visuales, fuese de fácil y amplio despliegue, cubriera más y de mejor manera necesidades específicas de nuestro contexto y tuviera en cuenta los aspectos mejorables de otras aplicaciones existentes.

2. Materiales y Métodos

Previamente, se analizaron las bases de este tipo de aplicaciones [11, 12] definiéndose los requerimientos básicos, así como las facilidades adicionales que enriquecieran e hicieran más funcional la aplicación. Se tuvieron en cuenta por igual las funcionalidades de algunas de las aplicaciones existentes. Varios requerimientos deben cumplirse en una aplicación móvil de esta índole:

- Una aplicación móvil destinada a personas con discapacidad visual requiere de una interactividad sencilla, que la navegación por sus funcionalidades y opciones sea de manera simple y directa.
- Un usuario espera que una aplicación de asistencia sea precisa, y genere poca o ninguna incertidumbre en sus operaciones y respuestas.
- Este tipo de aplicación posee necesidades de inmediatez. Por ejemplo, el caso de pagos requiere de rápidas verificaciones de los billetes.
- Teniendo en cuenta las limitantes de los usuarios de estas aplicaciones, se requiere una realimentación de las acciones realizadas, las operaciones ejecutadas por la aplicación y las respuestas que esta entrega, así como ayudas, de maneras sonoras (voz sintética o grabada, pitidos), y apticas (vibración)

Con esto se desarrolló una aplicación para dispositivos móviles, nombrada “aiSISTA”, para hacer accesible a las personas con discapacidades visuales el reconocimiento automático de billetes cubanos, de euro, dólar y tarjetas de diferentes tipos usadas en Cuba, empleando los procedimientos antes descritos. Esta nueva aplicación amplía las capacidades de las aplicaciones anteriores desarrolladas. Además del reconocimiento de billetes y tarjetas se incluyeron una serie de funcionalidades avanzadas adicionales. Las formas de interacción también se ampliaron, incluyéndose comandos de voz.

2.1 Desarrollo para Android

Android es un sistema operativo lanzado en 2007, basado en una versión modificada del núcleo del sistema operativo Linux y otros softwares de código abierto. Está diseñado primariamente para

dispositivos móviles con pantalla táctil, como teléfonos inteligentes y tabletas. Es desarrollado por un consorcio conocido como Open Handset Alliance y con el respaldo comercial de Google. En datos recabados en internet en agosto de 2024, a nivel mundial Android alcanzaba una cuota de mercado de un 71% [13]. En el caso de Cuba este valor alcanzaba 92.3% [14] lo que se corresponde con los datos publicados por ETECSA, el operador nacional de telecomunicaciones, donde ya en 2019 nueve (93%) de los diez fabricantes de móviles más usados en la red móvil del país usaban Android como sistema operativo [15]. Android es una plataforma con un menor coste general, más asequible el segmento poblacional de los usuarios de la aplicación, lo que lo hace aún más mayoritario entre ellos. Finalmente, la plataforma, da muchas facilidades de desarrollo y despliegue de las aplicaciones.

Como entorno integrado de desarrollo se empleó Android Studio, con el código funcional escrito en lenguaje de programación Java. Para el procesamiento de imagen y reconocimiento de patrones se emplea la biblioteca OpenCV en su versión 4.10, en compilación para Android y usando su API Java. Emplear OpenCV permite tener disponible en una única biblioteca funciones optimizadas para la mayoría de las operaciones primarias de procesamiento de imagen e inteligencia artificial a realizar. Para la ejecución se requiere como mínimo un dispositivo con CPU de arquitectura ARMv8 y versión 5.1 de Android. Una prioridad ha sido la optimización general, para reducir consumo de recursos y energía, y la entrega de resultados con baja latencia, aún en dispositivos de menores prestaciones o más antiguos.

2.2 Diseño

El diseño centrado en el usuario es un componente crítico de la interacción humano-computadora para personas con discapacidades visuales. Implica diseñar interfaces de usuario que sean intuitivas y fáciles de usar, y que tengan en cuenta las necesidades y preferencias de este tipo de usuarios [16, 17, 18]. Esto podría incluir características como texto grande y de alto contraste, menús y navegación bien organizados y comentarios de audio claros. Mediante una recopilación de información al respecto presente en otros trabajos se definieron las bases comunes del diseño, las consideradas innecesarias o redundantes, así como algunas nuevas.

La interfaz gráfica de la aplicación es de una simplificación máxima. No hay botones, ni letreros ni visualización innecesaria de la cámara (algo que consume energía). Si bien por las características de sus usuarios la pantalla se puede mantener en negro, aquí muestra la simbología de la función activa, por una razón más bien estética.

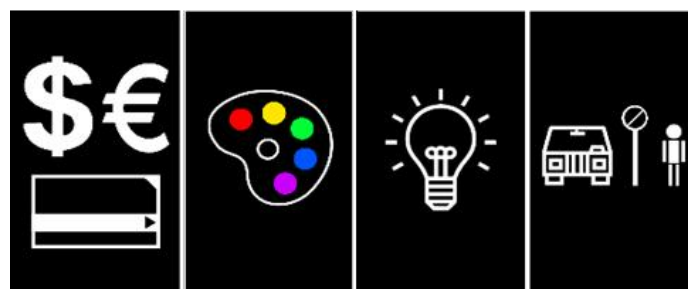


Fig.1 Simbología de algunas de las funciones.

2.3 Funcionalidades desarrolladas

La aplicación es capaz de realizar el reconocimiento de varias monedas y sus denominaciones, de diferentes tarjetas magnéticas, de diferentes colores, niveles de iluminación, objetos y su distancia estimada, textos/palabras, y transcribir dictados. En el desarrollo de estas funcionalidades se han empleado técnicas de procesamiento de imagen, de reconocimiento de patrones, estadísticas, y de inteligencia artificial basadas en redes neuronales.

Como preprocesamiento durante la captura y para contrarrestar las posibles condiciones de baja iluminación y sombras, siempre que sea necesario se activa automáticamente la iluminación del dispositivo. De las imágenes se estima su calidad, analizando los contrastes y el enfoque. Las imágenes que superan un umbral de calidad pasan a ser utilizadas, las que no se desechan.

2.3.1 Reconocer billetes y tarjetas

La aplicación provee reconocimiento de billetes de peso, dólar y euro, así como de las tarjetas de identificación y magnéticas empleadas en Cuba. Se siguen pautas similares a otros estudios [19].

Las imágenes captadas por la cámara se pasan a un clasificador basado en una red neuronal. La denominación con la mejor “respuesta” de esta red pasa por un proceso adicional de votación ponderada, y se devuelve la estimada como correcta, la que se anuncia al usuario y representa en pantalla el número/nombre en caracteres grandes contrastados.



Fig.2 Interacción durante el reconocimiento

El modelo de red neuronal empleada es MobileNet V3 [20] en su variante “pequeña”, con una capa de entrada de 224x224 píxeles. Se tomó el modelo pre-entrenado en el conjunto de datos ImageNet1K y se le sustituyó la sección de clasificación por uno ajustado al nuevo contexto, con 64 categorías de salida. Luego se aplicó un reentrenamiento mediante “transferencia de conocimiento” y “refinado”, en dos etapas por separado, con el conjunto de datos propios. Este *dataset* empleado contiene más de 45000 imágenes, reales y sintetizadas, en diferentes vistas, manipulación, condiciones de iluminación y entornos, de ambos lados de los billetes y tarjetas. Durante el entrenamiento el dataset se divide aleatoriamente en dos subconjuntos, de entrenamiento y de validación, en una proporción seleccionada de 80% y 20% respectivamente.

Durante el entrenamiento, y en cada iteración, las muestras recibieron alteraciones aleatorias adicionales (brillo, contraste, saturación, rotación, perspectiva, ampliación, enfoque, etc) para aumentar la diversidad (*augmentation*). Otros parámetros ajustados fueron la velocidad de aprendizaje (*learning rate*), su forma de variación (*decay*) donde se empleó el tipo *Cosine Annealing*, y el tamaño de los subconjuntos de muestras (*batch size*). Como las clases presentan desbalances en la cantidad de muestras se empleó FocalLoss [21]. Para el control de la calidad del entrenamiento y como regla de parada temprana del entrenamiento se tuvo en cuenta el comportamiento de las métricas de exactitud (acuracy), error (*loss*) y F1 (*F-score*) en ambos subconjuntos del dataset. PyTorch fue el framework de IA seleccionado, con aceleración mediante DirectML en un GPU AMD Radeon RX 470. La precisión y métrica F1 finales obtenidas sobre el conjunto de validación es de **99.19%**, con un error (*loss*) de **0.017**.

En cuanto al rendimiento, en el hardware empleado como línea base, un móvil Huawei Honor 5X (2015), con Android 6.0.1 y un CPU formado por 4 núcleos Cortex-A53 a 1.2 GHz y 4 Cortex-A53 a 1.5 GHz, durante la operación de inferencia, la latencia total (preprocesamiento + inferencia) es de unos 100 ms.

2.3.2 Reconocer objetos y escena

El reconocimiento de objetos está basado en un modelo red neuronal YOLOv8nano [22]. Permite detectar entidades de unas 80 clases. Para evitar confusiones, se eliminaron las más improbables en la vida cotidiana.

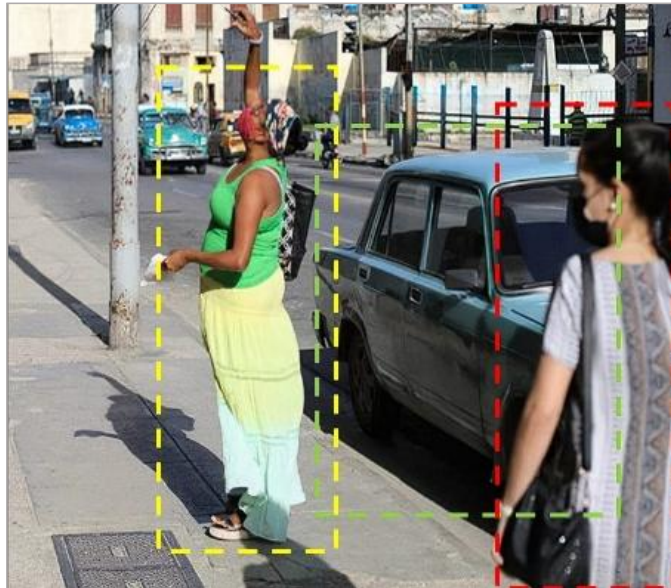


Fig.3 Ejemplo de detección. Esto no se muestra en pantalla

Las imágenes captadas pasan a la red neuronal, que devuelve un conjunto de regiones, sus clases estimadas y niveles de confianza para cada una. Las detecciones que superen un umbral de confianza pasan a una estimación de la distancia a que se encuentra la entidad detectada. Esto se realiza mediante operaciones trigonométricas de odometría visual que emplean parámetros intrínsecos del sistema óptico del dispositivo como la distancia focal y una “altura media” del objeto. Se eligen hasta tres detecciones prioritarias más cercanas, con distancia estimada inferior a cinco metros, de las que se localiza su posición, dígame a la izquierda, frente o derecha. También pueden localizarse arriba/abajo, pero se ha evitado para no abrumar al usuario y ganar tiempo.

Con un orden de prioridad, la aplicación dice lo detectado, la distancia estimada y la posición relativa. Por ejemplo, tomando el caso mostrado en la figura 3, la aplicación diría: “*persona, derecha, dos, persona, frente, cuatro, auto, derecha, cinco*”.

La alerta sobre los objetos cercanos, su posición y distancia pueden permitir al usuario navegar su entorno con algo más de seguridad. Por ejemplo, manteniendo el dispositivo con su cámara hacia el frente puede caminar mientras le va diciendo los posibles obstáculos a su paso.

La “descripción de escena” se realiza a partir de la detección de objetos anterior, pero sin límite en cuanto a la cantidad o distancia máxima. Con las clases determinadas para todas las detecciones válidas se elabora una sentencia descriptiva, que tiene en cuenta prioridades, número y género de las detecciones. Un número inferior o igual a tres detecciones de una misma clase se anuncia en su cantidad específica. Un número mayor que tres, se anuncia como “*varios*” o “*varias*” según el género. De igual manera ocurre con una sola detección de una clase, se construye como “*un/uno/una/unos/unas*”. La distancia y la posición no se tienen en cuenta en este caso. Tomando la figura 3, se describiría la escena como: “*dos personas y varios autos*”.

2.3.3 Navegación por marcadores.

Se emplean marcadores fiduciaros de alta densidad que contienen códigos numéricos y de verificación. Se pueden crear más de 16 millones de códigos diferentes. Estos códigos son índices a información previamente especificada (hasta mil entradas reservadas) o especificada a pedido, que se incluye en una base de datos local. Estos marcadores son impresos y adosados a los objetos o lugares que deben describir, para ser detectados y decodificados por la aplicación. Los colores han sido seleccionados para condensar información a 2 bits/cuadrícula [23], reducir las ambigüedades y maximizar el ahorro de tinta en la impresión. La estructura de la etiqueta también codifica su tamaño, que puede ser en tres medidas diferentes, 5, 10 y 20 centímetros de lado.

El detector está basado en [24] y emplea un descriptor entrenado con varios miles de muestras sintetizadas. Cuando se detecta una etiqueta se le indica al usuario su posición mediante voz sintetizada. Empleando odometría visual, junto a la posición se indica la distancia estimada a la etiqueta. En buenas condiciones, la etiqueta de mayor tamaño puede ser detectada y decodificada a diez metros de distancia.

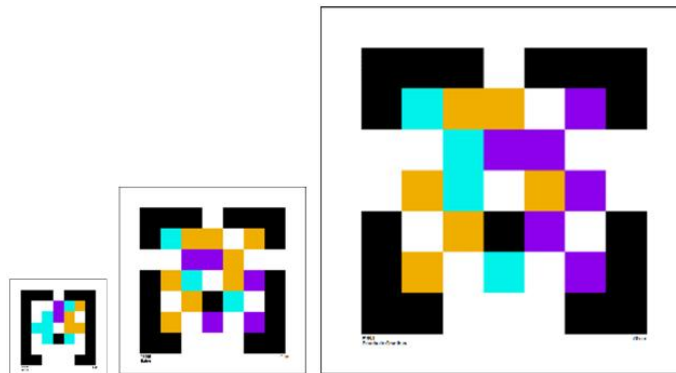


Fig.4 Ejemplos de marcadores para navegación, de 5, 10 y 20 centímetros

Para decodificar una etiqueta su imagen pasa primero por correcciones geométricas [25] y fotométricas. Los colores de cada casilla se comparan con los de referencia [26] y se extrae un código, que se valida con su CRC16, también incluido en la etiqueta. Si el CRC del código y el incluido son iguales se da por correctamente decodificado el marcador. Una vez es posible decodificar y validar la estructura y código del marcador, se anuncia la información referenciada en la base de datos.

Pueden generarse más de 16 millones de marcadores diferentes, de los que se han reservados unos mil para las necesidades comunes. Ejemplos son: “*peligro*”, “*baños*”, “*escaleras*”, “*parada de ómnibus*”, “*azúcar*”. Estas etiquetas comunes son publicadas para los usuarios junto a la aplicación. El resto puede generarse a pedido para ser empleadas en diferentes contextos.

2.3.4 Reconocer texto

La aplicación es capaz de detectar textos [27] de tipografía variada dispersos en el entorno y reconocerlo. Las regiones detectadas que cumplan cierto nivel mínimo de certidumbre, pasan al reconocimiento como tal del texto mediante la biblioteca Tesseract [28].

Para evitar confusiones o abrumar al usuario con una larga narración, actualmente se limita el reconocimiento hasta unas seis palabras, las más largas, de una longitud entre 3 y 14 caracteres. Las palabras reconocidas son dichas por el sintetizador de voz.

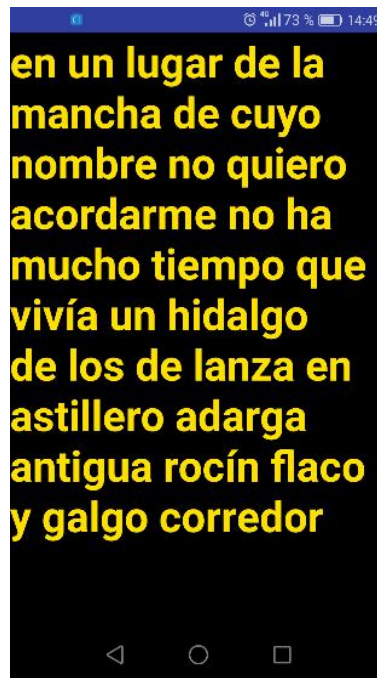


Fig.5 Transcripción de un dictado.

2.3.5 Transcribir dictados

La función de transcripción de dictados y de comandos de voz (ver más adelante), emplea el toolkit VOSK [29]. Esta funcionalidad puede ser útil para las personas que tienen dificultades para escribir en un teclado de pantalla táctil.

Mientras el usuario mantiene activa la función y habla de manera clara, se realiza el reconocimiento de voz y convierte lo dicho en un texto que va apareciendo transcrito en pantalla. Al finalizar de hablar, detener la operación o llegar al límite establecido, el texto se copia a memoria y se repite por la voz sintetizada para que el usuario pueda verificar que sea correcto. El texto copiado puede ser empleado en mensajes de texto, correos o, al mantenerse mostrado en pantalla, ser una vía de comunicación con una persona sorda.

2.3.6 Reconocer colores

Para reducir la influencia de partes de la imagen que no esté centrada en los objetos o regiones de interés se toma el 75% central de la imagen captada. El color en las imágenes captadas por la cámara puede sufrir corrimientos debido a problemas como el color de la iluminación, la especularidad de las superficies, la calidad de la cámara y como esta reacciona al espectro luminoso. Para reducir el efecto de estos corrimientos se aplican hasta dos correcciones automáticas de color o balance de blanco: el propio del dispositivo si dispone de ello, y otro por software. Estas correcciones reducen las desviaciones del color real de la escena y así los errores en el reconocimiento.

Creando un histograma de tono/saturación de los colores, se localiza en la imagen el color con mayor presencia, es decir, el más abundante y se compara con unos rangos preestablecidos para determinar los siguientes colores: rojo, naranja, amarillo, verde, azul, púrpura, blanco, gris y negro.

2.3.7 Estimar nivel de iluminación

Para el reconocimiento del nivel de iluminación se emplea el sensor especializado incluido en la mayoría de los móviles modernos. Este sensor se encuentra usualmente en la parte frontal de los dispositivos móviles y tiende a ser bastante direccional, es decir, responde más a la luz que le llega

en un ángulo estrecho. Los valores del nivel de iluminación se entregan en lux [30], una unidad derivada del Sistema Internacional de Unidades que define el flujo luminoso por unidad de área. Con una calibración experimental se fijaron varios rangos de niveles de iluminación: desde oscuro (4 lux o menos), muy baja, baja, media, alta, a muy alta (1000 lux o más). El valor de la iluminación captada se compara con estos rangos y se entrega el resultado mediante la voz sintetizada.

2.3.7 Ampliación de imagen

Para las personas con baja visión se incluye una función de ampliación de imagen con diferentes modos de representación.



Fig.6 Ampliación de imagen.

La ampliación de la imagen es al doble de la captada por la cámara del dispositivo. Se mostrarse en diferentes modos: imagen real, inversa, alto contraste, alto contraste inverso y alto contraste amarillo-negro.

2.3.8 Detección y lectura de QR

Esta funcionalidad realiza la detección y lectura de códigos QR. De manera similar a la navegación por etiquetas, guía al usuario indicando su posición, pero no la distancia. Si se detecta más de una, se reporta la de mayor tamaño en la imagen. Una vez se logra la decodificación correcta de la etiqueta, se dice su contenido y copia a memoria.

2.4 Interactividad y Realimentación

Se tiene en cuenta cómo los usuarios con discapacidad visual interactúan con un sistema y se centran en sus necesidades específicas para garantizar que los sistemas que diseñan sean accesibles y fáciles de usar. Esto implica realizar pruebas de usuario con participantes con discapacidad visual para identificar formas de mejorar el diseño de un sistema, como proporcionar mejores comentarios o utilizar métodos de entrada alternativos. Para un análisis más amplio del aspecto de la interactividad de la aplicación, ver [31].

2.4.1 Interacción manual

La mayor parte de la interacción con la aplicación y sus funcionalidades es de manera manual, siguiendo en lo posible las formas de manipulación a que están acostumbrados los usuarios objetivo al emplear otras aplicaciones de asistencia.

El usuario puede realizar deslizamientos en pantalla hacia los lados para acceder a las funciones de reconocimiento, o hacia arriba y abajo para información y ayuda. La función de cada sección es declarada mediante voz sintetizada por el dispositivo, guiando al usuario invidente. La ayuda

completa e información adicional está en una sección separada. Las secciones enlazan unas con otras en un ciclo de navegación.

La activación de las funciones se realiza mediante un toque continuo en la pantalla del dispositivo, que es reconocido con una vibración y con tonos cortos espaciados mientras está activada la función.

2.4.2 Interacción por comandos de voz

También empleando VOSK la aplicación adquiere capacidad de reconocer y ejecutar comandos de voz para cambiar entre funciones y/o activarlas, o pedir información. Con esto permite a las personas discapacitadas visuales trabajar con la aplicación sin necesidad de interactuar manualmente. También provee varias opciones adicionales solo accesibles mediante los comandos de voz. Esta capacidad se puede mantener activa aún esté la aplicación en un segundo plano o el dispositivo se encuentre con la pantalla apagada, lo que permite dar comandos sin necesidad de manipular el dispositivo.

Los comandos de voz están compuestos de una palabra de activación, y una o varias palabras de mando. Las palabras de activación de comando de voz son en este caso variantes de “asistencia”, por ejemplo, asistencia, asiste, asistir, asista. Si la aplicación reconoce que se ha dicho una palabra de activación, emite unos tonos indicando que esta lista para recibir el comando. Algunos ejemplos de palabras de mando establecidas son:

“**reconocer pesos**”, cambia a la opción de reconocimiento de billetes de moneda cubana.

“**activar pesos**”, cambia a esta opción, pero adicionalmente la activa.

“**hora**”, “**fecha**”, “**batería**”, la aplicación dice el estado de estos parámetros.

“**minimizar**”, lleva a segundo plano la aplicación.

La aplicación espera solo un corto tiempo por las palabras de comando. Si en este tiempo no se le ha dicho algo reconocible, emite unos tonos indicando que ha finalizado la espera, y el usuario necesitara decir de nuevo una palabra de activación.

2.4.2 Realimentación

Los usuarios con discapacidad visual necesitan una realimentación de las interacciones que realizan en el dispositivo y respuestas a las acciones de una manera consistente y fácilmente reconocible. Esta realimentación se realiza fundamentalmente de manera sonora, lo que incluye la síntesis de voz, de manera táctil, e incluso lumínica para las personas de baja visión.

Mientras están activadas en uso algunas operaciones, la aplicación emite tonos pausados de manera continua, así el usuario invidente se mantiene alerta sobre su actividad. Otros tonos anuncian el inicio y fin de las operaciones.

Un sistema Texto-a-Habla (TTS, Text to Speech) permite a las personas con discapacidad visual interactuar con dispositivos como computadoras o móviles. En 2009, la versión 1.6 de Android introdujo características de asistencia, donde destaca la capacidad TTS, con modelos y vocabularios incorporados a los dispositivos por los fabricantes o descargables. También pueden incorporarse sintetizadores diferentes, según la preferencia de los usuarios del sistema. El uso de estos sintetizadores de voz permite que todas las guías y respuestas de la aplicación sea vocalizada por el motor TTS activo en el dispositivo.

La ayuda e información fundamental la aplicación la dice mediante el sistema de voz sintetizada. La información completa aparece en la manual de usuario y la web dedicada.

Si la aplicación entregara las respuestas tan pronto las obtiene, como puede ocurrir en dispositivos rápidos, estas se solaparían confusamente. Por esto se aplica una corta pausa entre respuestas.

Una recomendación cuando se emplea la aplicación como asistencia de navegación en exteriores o en entornos con cierto nivel de ruido, es usar audífonos transductores, que transmiten el sonido a través de vibraciones, sin tapar los oídos ni atenuar los sonidos ambientales, útiles y necesarios a las personas ciegas para orientarse y comprender el entorno.

2.5 Privacidad

Las consideraciones de seguridad y privacidad en las aplicaciones informáticas de todo tipo han tomado más relevancia que nunca. Los datos enviados a sistemas de terceros, siendo procesados y almacenados a veces indefinidamente fuera del control del usuario, plantean muchas problemáticas de privacidad.

Teniendo esto en cuenta, la aplicación desarrollada no realiza ningún acceso a datos ni servicios externos. Todo el software, los datos y el procesamiento se mantienen dentro de la aplicación y el dispositivo. Las imágenes captadas se mantienen dentro del dispositivo. De hecho, la aplicación no almacena imágenes, las captura y directamente las procesa.

El aspecto negativo es que la aplicación requiere contener todas las bibliotecas, modelos y recursos necesarios para sus funcionalidades, lo que implica un aumento del tamaño de la aplicación. También implica un consumo energético adicional durante la ejecución de algunas funcionalidades.

3. Resultados y Discusión

Las pruebas de usuarios son una parte importante del proceso de desarrollo de aplicaciones para personas con discapacidades visuales. Implica trabajar con estos usuarios para probar la interactividad, las funcionalidades, la reactividad a las respuestas, y obtener comentarios sobre cómo mejorar la accesibilidad y la usabilidad, además encontrar posibles errores. Estos comentarios se utilizan para realizar mejoras en el diseño de la interfaz de usuario, la interactividad y realizar correcciones de errores.

Las pruebas fueron realizadas por varios usuarios voluntarios, ciegos, de baja visión y videntes. Entre estos usuarios los había avezados en el uso de los dispositivos móviles mediante sistemas de asistencia y todos poseían instalados y activados en sus dispositivos diferentes aplicaciones lectoras de pantalla, lo que permitió comprobar la compatibilidad de los mecanismos de interacción táctil y de síntesis de voz. Se comprobó además la compatibilidad con diversos dispositivos, el más antiguo lanzado en 2014. Esto es importante pues asegura en la medida de lo posible encontrar problemas e incompatibilidades dentro de la gran variedad de dispositivos Android y sus fabricantes, que se deben corregir. También se cubrió un amplio rango de potencias de cómputo para evaluar la inmediatez de las respuestas. Esto tiene en cuenta que en el sector poblacional de personas con discapacidades visuales tiende a tener menores recursos económicos lo que implica poseer dispositivos más antiguos o de menor prestación.

Los usuarios reportaron los problemas iniciales, con detalles como qué tipo de operación realizaban, que software adicional empleaban y las características de sus dispositivos. También realizaron recomendaciones de cambios y mejoras para mejor adaptación a sus necesidades y formas de uso de los dispositivos. Los problemas de incompatibilidades y conflictos se han ido resolviendo en lo fundamental en las versiones evolutivas de la aplicación.

La aplicación es puesta a disposición de los usuarios en una web dedicada, a través de artículos explicativos y enlaces de descarga, y mediante un canal en el sistema de mensajería Telegram al

que se adjunta un chat para realimentación con los usuarios, donde estos plantean e intercambian dudas, problemas, recomendaciones y soluciones. También se distribuye mediante Apkdis, la plataforma nacional de distribución de aplicaciones para Android. Quienes descargan e instalan la aplicación en sus dispositivos pueden a su vez distribuirla a otras personas sin limitaciones algunas. No requiere de registro, activación ni validación remota.

Los comentarios y opiniones de los usuarios recibidas por diferentes vías muestran una gran aceptación de la aplicación. No se ha hecho aún una evaluación general de aceptación, pero la aplicación anterior, más limitada en sus capacidades, alcanzó una evaluación de 4.6 sobre 5 en una encuesta respondida por 60 usuarios. Los intercambios con funcionarios de la ANCI y con usuarios permiten conocer que se ha logrado una amplia distribución de las aplicaciones previas entre los usuarios potenciales en todo el país, y esta no ha sido diferente. Al ayudar con el desenvolvimiento diario, cosa sensible y a veces complicada, los usuarios manifiestan agradecimiento al tenerseles en cuenta sus necesidades y que les faciliten la inclusión social e independencia.

4. Conclusiones

La aplicación aiSISTA se ha desarrollado para los discapacitados visuales como usuarios principales. Para ellos representa una ayuda a su independencia y desenvolvimiento de la cotidianeidad, así como una mejoría en su inserción y aporte social, ayudando en la inclusión de este segmento poblacional. Las opiniones han sido muy buenas y se ha mantenido una continua realimentación, con reportes de problemas, recomendaciones y soluciones, con los desarrolladores y entre los usuarios. La aplicación muestra una buena efectividad en las operaciones de reconocimiento si se siguen las pautas de manipulación. La compatibilidad está mayormente asegurada con dispositivos lanzados en los últimos 8 años. La interactividad está hecha a la medida de las necesidades y posibilidades de las personas ciegas o de baja visión, facilitando su uso y la comprensión de las respuestas. aiSISTA es una herramienta en evolución y varios aspectos se han tenido en cuenta para su desarrollo futuro, mejorar sus capacidades actuales e incluir nuevas, tales como: reconocer textos más amplios y en documentos, mejor descripción de escena, emplear un detector de etiquetas más robusto, emplear el sensor de aceleración para alertar ante posibles colisiones durante la marcha, realizar localización y navegación empleando GPS e información geográfica, facilitar la localización y el reconocimiento de códigos QR (ya realizado), incluir el reconocimiento y descripción de personas, incluir nuevos comandos de voz con más acciones, incluso extender a otras áreas de asistencia y discapacidades, como el reconocimiento de lenguajes de señas. En general, aiSISTA se convierte en una poderosa herramienta para las personas discapacitadas visuales al proveerles con la habilidad de identificar e interactuar con su entorno en maneras que antes les era muy difícil o imposible.

Referencias

1. World Health Organization, "Factsheet on visual impairment and blindness, world health organization", *World Blind Union*, 2012
2. World Health Organization. "World report on vision. Executive Summary". *Department of Non communicable Diseases*, 2019
3. Rodríguez, A., et al. "Open Challenges of Blind People Using Smartphones". *International Journal of Human-Computer Interaction*, 36:17, 1605-1622, 2020
4. W3C. Web Content Accessibility Guidelines (WCAG) 2.1. *W3C Recommendation 05 June 2018*. <https://www.w3.org/TR/WCAG21/>
5. Augustin, A. et al. "Anxiety and depression prevalence rates in age-related macular degeneration". *Investigative Ophthalmology and Vision Science*, 48, 1498-1503, 2007

6. Sagale, A. A., Chaudhari, A. "Assistive Technologies for Visually Impaired Users on Android Mobile Phones". *International Journal of Computer Applications*, Vol. 182 No. 32, 2018
7. Dobosz, K. "Designing Mobile Applications for Visually Impaired People". *Visually Impaired: Assistive Technologies, Challenges and Coping Strategies*, pp.103-126, Nova Science Pub, 2017
8. Leo, M., Medioni, G., Trivedi, M., Kanade, T., Farinella, G.M. "Computer vision for assistive technologies". *Computer Vision and Image Understanding* 154, 1–15 , 2016
9. Salas, M. "Asistencia a discapacitados visuales para reconocer billetes mediante aplicación móvil". *Revista Cubana de Ciencias Informáticas*, Vol. 15, No. Especial UCIENCIA II, septiembre, 2021, ISSN: 2227-1899, RNPS: 2301
10. Salas, M. "qBillete. Asistencia en dispositivos móviles a personas con discapacidades visuales".
11. Rashid, A., Prati, A., Cucchiara, R. "On the design of embedded solutions to banknote recognition". *Optical Engineering* 52(9), pp. 93-106. 2013
12. Solymar, Z., Stubendek, A., Radvanyi, M., Karacs, K. "Banknote recognition for visually impaired". *20th European Conference on Circuit Theory and Design*, 2020
13. *Memorias de la Convención y Feria Informática 2022*. InfoTIC. ISBN 978-959-7255-02-4, 2022
14. Statcounter MobileOS Global. <https://gs.statcounter.com/os-market-share/mobile/> , 2024
15. Statcounter MobileOS Cuba. <https://gs.statcounter.com/os-market-share/mobile/cuba> , 2024
16. Cubadebate. "¿Qué teléfonos móviles se usan más en Cuba?" www.cubadebate.cu/noticias/2020/03/16/que-telefonos-moviles-se-usan-mas-en-cuba-infografia-y-top-10/ , 2020
17. Rogers, J., Sharp, Y., Preece, H. *Interaction design: beyond human computer interaction*, 6th Edition, Wiley, 2023.
18. Qureshi, H.H., Wong, D.H.T. "Requirements of a Mobile Application Design Model for Visually Impaired People". *Open International Journal of Informatics (OIJI)*, Vol. 7 Special Issue 2, 2019
19. Qureshi, H.H., Hooi-Ten Wong, D. "Usability of user-centric mobile application design from visually impaired people's perspective". *Universal access in human-computer interaction. Design approaches and supporting technologies*. HCII 2020. Springer
20. Solymar, Z., Stubendek, A., Radvanyi, M., Karacs, K. "Banknote recognition for visually impaired". *20th European Conference on Circuit Theory and Design*. 2020
21. Howard, A., Sandler, M. et al. "Searching for MobileNetV3". 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 1314-1324
22. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P. "Focal Loss for Dense Object Detection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.42, no.2, pp. 318-327, 2020
23. Jocher, G., Chaurasia, A., Qiu, J. "YOLOv8", *Ultralytics*, 2023
24. Querini, M., Grillo, A., Lentini, A., Italiano, G.F. "2D Color Barcodes for Mobile Phones". *International Journal of Computer Science and Applications*, Vol. 8 No. 1, pp. 136 - 155, 2011.
25. Viola, P., Jones, M. "Rapid Object Detection using a Boosted Cascade of Simple Features". *IEEE Conf. Comput. Visual Pattern Recognition*. 1.I-511.10.1109/CVPR.2001.990517. 2021
26. Evangelidis, G.D., Psarakis, E.Z. "Parametric Image Alignment Using Enhanced Correlation Coefficient Maximization". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30, 1858-1865. 2008
27. Querini, M., Italiano, G.F. "Reliability and Data Density in High Capacity Color Barcodes". *Computer Science and Information Systems* 11(4):1595–1615, 2014

28. Zhou, X. et al. "EAST: An Efficient and Accurate Scene Text Detector". arXiv:1704.03155 [cs.CV], 2017
29. Ooms, J. "tesseract: Open Source OCR Engine". *Github/ropensci/tesseract*, 2023
30. Nickolay, V. S., et al. "Vosk Speech Recognition Toolkit: Offline speech recognition API for Android, iOS, Raspberry Pi and servers with Python, Java, C# and Node", 2024
31. NIST, "The NIST Reference on Constants, Units and Uncertainty", 2023
32. Salas, M. aSISTA. Aplicación de Asistencia a Personas con Discapacidades Visuales en Dispositivos Móviles. *Interacción. Revista Digital de la AIPO*, Vol.4 Núm.1 pp. 29-39, 2023

Agradecimientos

Se agradece a la Asociación Nacional del Ciego de Cuba su colaboración y excelente recepción de estos desarrollos. A ellos se les dedica este trabajo.

Conflicto de Intereses

No existen conflictos de intereses.

Contribución de los autores

Maikel Salas Zaldívar. <https://orcid.org/0000-0001-9640-3258>