

Asistencia móvil inteligente para personas con discapacidad visual

Maikel Salas Zaldívar

RESUMEN

La discapacidad visual limita significativamente la interacción con el entorno, pero los avances en inteligencia artificial ofrecen nuevas oportunidades de autonomía. Aunque los dispositivos móviles no están diseñados originalmente para personas con ceguera o baja visión, muchos los utilizan mediante tecnologías de accesibilidad como lectores de pantalla. Estos dispositivos integran cámaras y sensores que, combinados con modelos de IA en el propio terminal, permiten interpretar el mundo visual de forma inteligente y en tiempo real. Se ha desarrollado una aplicación para Android que emplea visión por computadora, redes neuronales y reconocimiento de patrones para identificar billetes, tarjetas bancarias, objetos cotidianos, colores y niveles de iluminación, dictados orales, textos, marcadores, además de ayudar a comprender y navegar el entorno circundante. Todo el procesamiento se realiza localmente, garantizando privacidad y respuesta inmediata. La interacción se basa en gestos táctiles intuitivos, y la retroalimentación se entrega mediante voz sintetizada, sonidos o señales hápticas. Al integrar inteligencia artificial en el dispositivo móvil, esta solución transforma un smartphone en un asistente visual inteligente, potenciando la independencia, seguridad y participación social de las personas con discapacidad visual.

Palabras claves: aplicación, asistencia, discapacidad visual, inteligencia artificial, dispositivo móvil

ABSTRACT

Visual impairment significantly limits interaction with the environment, but advances in artificial intelligence (AI) offer new opportunities for autonomy. Although mobile devices were not originally designed for people who are blind or have low vision, many already use them through accessibility technologies such as screen readers. These devices incorporate cameras and sensors that, when combined with on-device AI models, enable intelligent, real-time interpretation of the visual world. An Android application has been developed that leverages computer vision, neural networks, and pattern recognition to identify banknotes, credit cards, everyday objects, colors, text, dictations, markers and lighting levels, as well as to understand and navigate the surrounding environment. All processing occurs locally on the device, ensuring user privacy and immediate responsiveness. Interaction is based on intuitive touch gestures, while feedback is delivered through synthesized speech, audio cues, or haptic signals. By embedding artificial intelligence into the mobile device, this solution transforms a smartphone into an intelligent visual assistant, enhancing independence, safety, and social inclusion for people with visual impairments.

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

Intelligent mobile assistance for people with visual impairments

1. -INTRODUCCIÓN

La visión constituye un sentido fundamental en la percepción e interacción del ser humano con su entorno, desempeñando un papel crítico en la realización de actividades cotidianas. Según estimaciones de la Organización Mundial de la Salud (OMS), en 2019 existían aproximadamente 285 millones de personas en el mundo con algún grado de discapacidad visual, más del 4% de la población global, de las cuales treinta y nueve millones eran ciegas [1]. En el contexto cubano, más de 30000 personas con ceguera o baja visión están afiliadas a la Asociación Nacional del Ciego (ANCI), lo que evidencia la relevancia del problema a nivel nacional.

Recibido: 11/2025 Aceptado: 02/2026

En las últimas décadas, los dispositivos móviles inteligentes se han consolidado como herramientas centrales de comunicación, acceso a la información y gestión de tareas diarias. No obstante, su diseño predominante, basado en interfaces visuales y táctiles, plantea barreras significativas para las personas con discapacidad visual, quienes enfrentan dificultades para interactuar con elementos que requieren localización espacial en la pantalla [2]. Aunque existen estándares internacionales de accesibilidad, como las Pautas de Accesibilidad al Contenido Web (WCAG) del Consorcio W3C [3], orientadas a garantizar la usabilidad de software y sitios web mediante criterios de contraste, navegación y compatibilidad con tecnologías asistivas, estas no resuelven por sí solas los desafíos inherentes a la percepción del entorno físico.

Las personas con discapacidad visual enfrentan diariamente situaciones que generan estrés físico y psicológico, afectando su calidad de vida [4]. Entre estas dificultades destacan: la orientación y comprensión del entorno más allá del alcance del bastón o la audición, la detección de obstáculos en la vía pública, la identificación de colores o niveles de iluminación, y el reconocimiento fiable de billetes, monedas o tarjetas bancarias. Si bien han surgido tecnologías asistivas, como lectores de pantalla, *displays* braille, sistemas de reconocimiento de voz y software especializado, que facilitan la interacción con dispositivos digitales [5] persiste una brecha crítica en la capacidad de interpretar información visual del mundo real.

Es en este escenario donde las herramientas inteligentes, basadas en inteligencia artificial (IA), visión por computadora y reconocimiento de patrones, adquieren un rol transformador. Aplicaciones como SeeingAI, Lookout, TapTapSee, CurrencyIdentifier, MoneyReader, ViaOptaDaily, Blind-Droid Wallet, EyeNote, entre otras [6 - 11], demuestran el potencial de estos enfoques para asistir en tareas como el reconocimiento de objetos, texto, moneda o navegación espacial. Sin embargo, la mayoría de estas soluciones presenta limitaciones significativas en contextos como el cubano, por ejemplo: algunas tienen dependencia exclusiva del sistema operativo iOS (plataforma de baja penetración en la isla), total ausencia de soporte para billetes, monedas y tarjetas nacionales, interfaces gráficas innecesarias para usuarios no videntes, y dependencia de conectividad a servicios en la nube, lo que implica riesgos en privacidad, costos y fallos de disponibilidad ante intermitencias.

Además, muchas de estas aplicaciones ofrecen funcionalidades fragmentadas, por ejemplo, una solo reconoce billetes, otra solo texto, obligando a los usuarios a gestionar múltiples herramientas con lógicas de interacción heterogéneas.

Ante estas limitaciones, surge la necesidad de desarrollar una aplicación inteligente unificada, diseñada para dispositivos móviles Android, plataforma dominante en Cuba, que integre de forma coherente múltiples capacidades asistivas mediante modelos de IA ejecutados localmente. Esta solución debe abordar necesidades del contexto nacional (como el reconocimiento de moneda y tarjetas cubanas), priorizar en su diseño la accesibilidad basada exclusivamente en retroalimentación auditiva y háptica, garantizar autonomía operativa sin dependencia de conectividad externa y superar las deficiencias observadas en soluciones existentes. El objetivo último es potenciar la autonomía, seguridad e inclusión social de las personas con discapacidad visual mediante el uso de herramientas inteligentes integradas en dispositivos de uso cotidiano.

2. -Análisis de Antecedentes

Previamente al desarrollo de la solución propuesta, se realizó un análisis exhaustivo de las bases conceptuales y técnicas que sustentan las aplicaciones móviles orientadas a personas con discapacidad visual, tomando como referentes estudios previos [12, 13]. Este análisis permitió definir tanto los requisitos funcionales y no funcionales esenciales como un conjunto de funcionalidades adicionales que enriquecieran la utilidad y experiencia de uso del sistema. Paralelamente, se evaluaron críticamente las capacidades de aplicaciones existentes en el mercado, con el fin de identificar buenas prácticas y limitaciones a superar. En concordancia con la literatura especializada [14, 15], se establecieron los siguientes requisitos fundamentales para una aplicación móvil de asistencia basada en inteligencia artificial dirigida a usuarios con discapacidad visual:

- Interacción simplificada: la navegación entre funcionalidades debe ser intuitiva, lineal, sin elementos visuales redundantes o distractores, privilegiando mecanismos como gestos táctiles o comandos de voz.
- Alta precisión y confiabilidad: el usuario requiere respuestas consistentes y certeras, especialmente en tareas críticas como la identificación monetaria, donde la incertidumbre puede generar consecuencias negativas.
- Inmediatez en la respuesta: ciertas operaciones, como la verificación de billetes durante transacciones comerciales, demandan tiempos de respuesta mínimos, lo que exige optimización algorítmica y, preferiblemente, ejecución local de modelos de inteligencia artificial.
- Retroalimentación multimodal: dada la imposibilidad o baja capacidad de percepción visual, toda acción, estado o resultado debe comunicarse mediante canales alternativos, tales como voz sintetizada o pregrabada, señales sonoras (pitidos codificados) y retroalimentación háptica (vibraciones con patrones significativos).

Con base en estos principios, se desarrolló inicialmente qBillete [16, 17], una aplicación móvil para el sistema operativo Android que emplea técnicas de procesamiento de imágenes para el reconocimiento automático de billetes de curso legal en Cuba (peso), así como divisas internacionales (euro y dólar estadounidense), y diversos tipos de tarjetas utilizadas en el

contexto nacional (bancarias, de identificación). La aplicación fue diseñada como una herramienta accesible, priorizando la autonomía del usuario y la ejecución eficiente en dispositivos de gama media, comunes en el entorno cubano.

Posteriormente, qBillete fue ampliada significativamente. Se actualizaron sus funcionalidades y se incorporaron otras nuevas, tales como detección de colores, estimación de niveles de iluminación ambiental, identificación de objetos cotidianos y asistencia básica para la orientación espacial, ahora en su mayoría empleando inteligencia artificial basada en modelos de redes neuronales. Asimismo, se enriquecieron los mecanismos de interacción mediante la integración de comandos de voz, permitiendo una operación más fluida y natural, alineada con los paradigmas actuales de interfaces inteligentes centradas en el usuario. Esta evolución se ha sido denominada aSista, una herramienta inteligente integral de asistencia, adaptada a las necesidades reales y al contexto tecnológico de las personas con discapacidad visual en Cuba, y es la que describe este trabajo.

3. -Métodos

3.1 Desarrollo para Android

Android constituye un sistema operativo de código abierto lanzado en 2007, basado en una versión modificada del núcleo Linux y complementado con una amplia gama de componentes de software libre. Diseñado específicamente para dispositivos móviles con interfaces táctiles, tales como teléfonos inteligentes y tabletas. Su arquitectura modular, flexibilidad y ecosistema robusto lo han posicionado como la plataforma dominante a nivel global. Según datos de StatCounter, en octubre de 2025 Android alcanzaba una cuota de mercado (en teléfonos móviles) del 75% a escala mundial. En el contexto cubano, esta cifra asciende al 95%, lo cual se alinea con reportes del operador nacional de telecomunicaciones ETECSA, que indicaba en 2020 que al menos nueve de cada diez teléfonos móviles utilizados en la red nacional empleaban Android como sistema operativo. Esta alta penetración en Cuba justifica su elección para el despliegue con enfoque nacional.

El procesamiento visual y de reconocimiento de patrones se implementó mediante la biblioteca OpenCV en su versión 4.10, compilada para Android, empleándose el lenguaje de programación Java para la implementación del código funcional.

La aplicación resultante se ejecuta en dispositivos con arquitectura de procesador ARMv8 y sistema operativo Android 5.1 (API 22), como requerimientos mínimos. Dado que el desempeño en tiempo real y la eficiencia energética son críticos para la usabilidad en contextos reales, especialmente en la gama de dispositivos comunes en Cuba, se priorizó rigurosamente la optimización algorítmica, computacional y funcional en todas las etapas del desarrollo. La aplicación contiene seis modelos diferentes de inteligencia artificial (descritos en sus acápitales) y sus funciones operan de forma autónoma, con baja latencia y sin conectividad externa, garantizando accesibilidad, privacidad y respuesta inmediata para el usuario.

3.2 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 [18, 19]. Esto podría incluir características como texto grande y de alto contraste, menús y navegación bien organizados y comentarios de audio claros [20]. Mediante una recopilación de información sobre otros trabajos se definieron las bases comunes del diseño, las consideradas innecesarias o redundantes, así como algunas noveles.

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 (Figura 1).

3.3 Funcionalidades desarrolladas

La aplicación implementa un conjunto de funcionalidades inteligentes que permiten el reconocimiento y análisis del entorno visual mediante técnicas avanzadas de procesamiento de señales e inteligencia artificial. Específicamente, es capaz de identificar múltiples tipos de monedas y sus respectivas denominaciones, distinguir diversas tarjetas magnéticas (bancarias, de transporte, de identificación, entre otras), reconocer colores, estimar niveles de iluminación ambiental, detectar y clasificar objetos cotidianos junto con una estimación aproximada de su distancia, extraer y transcribir texto impreso y transcribir dictados de voz mediante reconocimiento automático del habla.

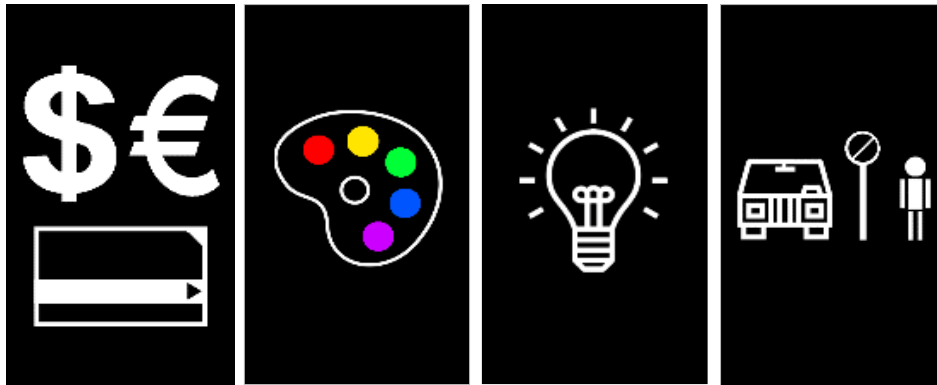


Figura 1

Simbología en pantalla del dispositivo de algunas de las funciones de la aplicación.

Reconocer billetes y tarjetas, reconocer color, estimar nivel de iluminación, detectar objetos y estimar distancia y posición.

El desarrollo de estas capacidades se fundamenta en una combinación de procesamiento digital de imágenes, reconocimiento de patrones, métodos estadísticos y modelos de inteligencia artificial basados en redes neuronales profundas. La ejecución es toda local, en el dispositivo, lo que garantiza baja latencia, autonomía operativa y preservación de la privacidad del usuario, aspectos críticos en contextos con limitaciones de conectividad, como el cubano.

Durante la fase de adquisición visual, se aplican estrategias de preprocesamiento inteligente para asegurar la calidad de los datos de entrada. En condiciones de baja iluminación o presencia de sombras que puedan comprometer la precisión del análisis, el sistema activa de forma automática la linterna del dispositivo, optimizando así las condiciones de captura sin intervención del usuario. Posteriormente, cada imagen capturada es sometida a una evaluación algorítmica de su calidad, mediante métricas objetivas que analizan parámetros como el contraste local, la nitidez (enfoque) y la presencia de ruido. Únicamente aquellas imágenes que superan un umbral predefinido de calidad son procesadas y las que no cumplen con este criterio son descartadas, evitando inferencias erróneas y mejorando la robustez general del sistema. Este enfoque basado en la calidad perceptual y la toma de decisiones autónoma constituye un componente esencial de la arquitectura de herramientas inteligentes orientadas a la asistencia confiable y eficiente de personas con discapacidad visual.

3.3.1 Reconocer billetes y tarjetas

La aplicación implementa un sistema inteligente de reconocimiento visual capaz de identificar billetes en peso cubano, dólar estadounidense y euro, así como tarjetas de identificación y tarjetas magnéticas utilizadas en Cuba (Figura 2). Las imágenes capturadas por la cámara del dispositivo son sometidas a un filtro de calidad; únicamente aquellas que cumplen con criterios mínimos de nitidez, contraste y enfoque se procesan. Esto se realiza para todas las funciones, no solo para el reconocimiento de billetes y tarjetas. Estas imágenes aceptables se alimentan a un clasificador basado en el modelo de red neuronal MobileNetV3-Small [21], seleccionado por su equilibrio entre precisión y eficiencia computacional en dispositivos móviles.

Partiendo de una versión del modelo pre-entrenado en el *dataset* ImageNet-1K, se reemplazó la capa de clasificación final y se realizó un refinamiento. El *dataset* propio incluye más de 45000 imágenes, combinando tomas reales y sintetizadas representando ambos lados de los billetes y tarjetas bajo variadas condiciones: ángulos, iluminación, oclusiones parciales y entornos diversos. Se incluyen además varios miles de muestras de posibles estructuras de fondo (*background*). Para incrementar la robustez, se aumentó más aun la variabilidad de los datos aplicando durante el entrenamiento transformaciones aleatorias del tipo que aparecen en el uso real (brillo, contraste, saturación, rotación, perspectiva y escala).

Dado el desbalance entre clases en el conjunto de datos, se empleó la función de pérdida Focal Loss [22] para mitigar el sesgo hacia categorías sobre-representadas. El entrenamiento se realizó con PyTorch, alcanzando el modelo final una precisión (*accuracy*) de 99.19 % y una métrica F1 de 99.19 % con una pérdida (*loss*) de 0.017, sobre el conjunto de validación.



Figura 2
Interacción durante el reconocimiento.

Tras la inferencia, la predicción con mayor activación pasa por un mecanismo de votación ponderada usando varios resultados previos acumulados, que refuerza la confiabilidad del resultado. La clase estimada como correcta se comunica al usuario mediante voz sintetizada y se muestra en pantalla con caracteres grandes y de alto contraste.

En términos de rendimiento, evaluado en un Huawei Honor 5X (lanzado en 2015) con Android 6.0.1, CPU octa-core (4×1.2 GHz + 4×1.5 GHz Cortex-A53), el sistema logra una latencia total de aproximadamente 100 ms, incluyendo preprocesamiento e inferencia, lo que permite una interacción en tiempo real.

3.3.2 Reconocer objetos y escena

El reconocimiento de objetos se basa en un modelo de red neuronal de arquitectura tipo YOLO, en este caso la versión *9tiny* (minúscula) [23]. Este modelo permite detectar instancias pertenecientes a un conjunto de 80 clases de objetos; sin embargo, con el objetivo de reducir ambigüedades y mejorar la relevancia contextual, se ignoran aquellas categorías poco probables en entornos cotidianos.

Las imágenes capturadas por la cámara del dispositivo se procesan en tiempo real mediante esta red, la cual genera un conjunto de regiones de interés, cada una asociada a una clase estimada y un nivel de confianza. Las detecciones que superan el umbral de aceptación avanzan a una etapa posterior de estimación de distancia mediante métodos de odometría visual basados en trigonometría, que emplean los parámetros intrínsecos de la cámara (tamaño del sensor, resolución, distancia focal) y una altura media para cada clase de objeto. A partir de este cálculo, se seleccionan hasta tres detecciones prioritarias ubicadas a menos de cinco metros, y se determina su posición relativa respecto al usuario: izquierda, frente o derecha. Aunque el sistema es capaz de estimar también posiciones verticales (arriba/abajo), esta dimensión se ha desactivado intencionalmente para evitar sobrecarga cognitiva y garantizar respuestas ágiles. La retroalimentación se entrega mediante voz sintetizada, siguiendo un orden de prioridad (proximidad y relevancia), e incluye la clase del objeto, su ubicación horizontal y su distancia estimada (en metros). Por ejemplo, ante una escena como la de la Figura 3, la aplicación anunciaría: “persona, derecha, dos; persona, frente, cuatro; auto, derecha, cinco”.

Esta funcionalidad permite al usuario navegar con mayor seguridad, especialmente al mantener el dispositivo orientado hacia el frente mientras camina, recibiendo alertas en tiempo real sobre obstáculos o elementos relevantes en su trayectoria.

Complementariamente, la descripción global de la escena se construye a partir del mismo conjunto de detecciones, pero sin restricciones de distancia ni número máximo de objetos. A partir de las clases válidas detectadas, se genera una oración descriptiva coherente, que aplica reglas lingüísticas para concordancia de número y género. Cuando hay tres o menos instancias de una misma clase, se especifica la cantidad exacta (ejemplo: “dos personas”); si hay más de tres, se utiliza el cuantificador “varios” o “varias” según el género gramatical de la clase (ejemplo: “varios autos”). En el caso de una sola instancia, se emplea el artículo indefinido adecuado (“un”, “una”, etc.). A diferencia del modo de navegación, no se incluyen distancia ni posición en esta descripción. Así, la escena de la Figura 3 se resumiría como: “dos personas y varios autos”.

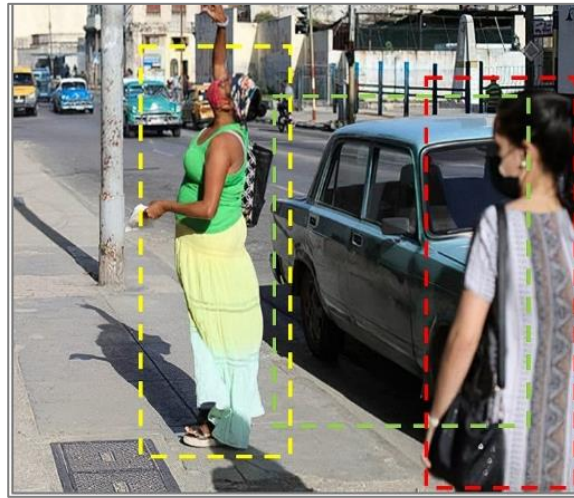


Figura 3
Detección en escena. Esto no se muestra en pantalla.

3.3.3 Navegación por marcadores

La aplicación incorpora un sistema de marcadores fiduciaros de alta densidad, diseñados para codificar información contextual de forma robusta y eficiente. Cada marcador contiene un código numérico único de 24 bits y un código de verificación CRC16, permitiendo la generación de más de 16 millones de combinaciones distintas. Estos códigos actúan como índices que referencian entradas en una base de datos local, de los cuales, hasta 1000 son registros predefinidos (por ejemplo, “peligro”, “baños”, “escaleras”, “parada de ómnibus”, “azúcar”). Los marcadores se imprimen y adosan físicamente a objetos o ubicaciones relevantes, facilitando su identificación mediante la cámara del dispositivo. Otros marcadores pueden ser creados y añadidos a la aplicación por los autores, bajo solicitud de usuarios.

El diseño del marcador optimiza la densidad de datos y la eficiencia de impresión: emplea un esquema de codificación por color que almacena 2 bits por celda [24], seleccionando tonos que minimizan ambigüedades cromáticas bajo diversas condiciones de iluminación y reducen el consumo de tinta en la impresión (Figura 4). Su estructura facilita la detección y su geometría codifica su tamaño físico, lo que permite una estimación más precisa de la distancia al ser detectado el marcador.

El detector se basa en [25], y fue entrenado con miles de muestras sintetizadas que simulan variaciones de perspectiva, escala, iluminación y ruido. Al detectar un marcador, se obtiene su posición en la imagen y estima su distancia mediante odometría visual, utilizando los parámetros intrínsecos de la cámara y la escala conocida del marcador, luego el sistema informa su presencia y posición relativa (frente, arriba/abajo, derecha/izquierda y distancia en metros) mediante voz sintetizada.



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

El proceso de decodificación incluye una etapa de corrección geométrica [26] y normalización fotométrica para compensar distorsiones y variaciones de iluminación. Posteriormente, los colores de cada celda se comparan con una paleta de referencia para extraer el código binario subyacente [27]. Este código se valida mediante su CRC16 embebido; solo si el valor calculado del CRC del dato decodificado coincide con el almacenado en el marcador, se considera una decodificación exitosa. Una vez validado, el sistema recupera la descripción asociada desde la base de datos local y la anuncia al usuario mediante voz. En condiciones óptimas (iluminación, posición, calidad del sensor de imagen), los marcadores de mayor tamaño pueden ser detectados y decodificados a distancias de hasta 10 metros.

Los marcadores predefinidos de uso común (como “peligro”, “baños”, “escaleras”, “parada de ómnibus”, “azúcar”), se publican en la página web oficial de la aplicación, permitiendo su descarga e impresión por parte de la comunidad, lo que fomenta la creación colectiva de entornos más accesibles.

3.3.4 Reconocer texto

La aplicación integra una funcionalidad para la detección y reconocimiento de texto en escenas naturales, capaz de identificar tipografías variadas distribuidas en el entorno. El modelo EAST de detección [28] localiza regiones potencialmente textuales en la imagen capturada (Figura 5). Aquellas regiones que superan un umbral mínimo de certidumbre son enviadas al módulo de reconocimiento óptico de caracteres (OCR), implementado mediante el modelo DenseNet-CTC [29], capaz de lidiar con tipografía diversa y compleja, y como *fallback*, con la biblioteca Tesseract [30] con un modelo específico para español, aunque sus resultados no son mucho mejores, al emplear esta biblioteca un modelo estándar orientado a tipos comunes.

Las palabras se ordenan por su posición, de izquierda a derecha y de arriba abajo. Con el objetivo de preservar la usabilidad y evitar sobrecarga del usuario, el sistema limita la salida a un máximo de seis palabras, cada una con una longitud comprendida entre 3 y 14 caracteres. Este criterio prioriza la claridad y brevedad de la retroalimentación. Las palabras reconocidas se comunican al usuario exclusivamente mediante síntesis de voz, garantizando una interacción accesible y eficiente en contextos de movilidad o exploración ambiental.



Figura 5

Detección de regiones textuales en la imagen. Esto no se muestra en la aplicación.

3.3.5 Transcribir dictados

La función de transcripción de dictados y comandos de voz se implementa mediante el *toolkit* de código abierto VOSK [31], una solución de reconocimiento de voz para dispositivos con recursos limitados. Esta herramienta convierte el habla del usuario en texto en tiempo real, hasta una longitud predefinida, constituyendo una alternativa accesible para personas con dificultades para interactuar con teclados táctiles.

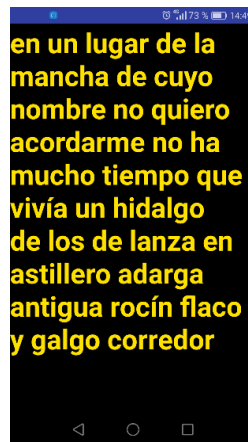


Figura 6

Transcripción de un dictado. El texto se copia a memoria y se mantiene visible en pantalla.

Mientras la función permanece activa y el usuario habla con claridad, el texto transcrito se muestra progresivamente en pantalla (Figura 6). Al finalizar la entrada, ya sea por pausa prolongada, interrupción explícita o al alcanzar el límite de caracteres, el sistema copia el texto a memoria, lo reproduce mediante síntesis de voz para su verificación auditiva y lo mantiene visible. Esta doble retroalimentación (auditiva y visual) permite al usuario confirmar la precisión de la transcripción. El texto resultante puede emplearse posteriormente en mensajes de texto, correos electrónicos o, al mantenerse en pantalla, servir como medio de comunicación con personas sordas, ampliando así su utilidad más allá de la interacción individual.

3.3.6 Reconocer colores

Con el fin de minimizar la influencia de áreas periféricas irrelevantes en el análisis visual, se procesa únicamente el 75 % central de la imagen capturada. Adicionalmente, dado que los colores pueden verse distorsionados por factores la iluminación local, la especularidad de las superficies, la calidad del sensor o la respuesta espectral de la cámara, se aplican hasta dos etapas sucesivas de corrección automática de color: primero, el balance de blancos nativo del dispositivo (si está disponible), y luego un algoritmo de corrección por software. Estas medidas reducen las desviaciones cromáticas respecto a la escena real. Para la identificación de colores, se construye un histograma en el espacio de tono y saturación y se determina el color dominante en la imagen. Este valor se compara con rangos predefinidos de nueve categorías: rojo, naranja, amarillo, verde, azul, púrpura, blanco, gris o negro. Su precisión se reduce ante ambigüedades provocadas por el material/superficie o la iluminación local.

3.3.7 Estimar nivel de iluminación

En cuanto al nivel de iluminación ambiental, la aplicación aprovecha el sensor de luz ambiental presente en la mayoría de los dispositivos móviles modernos, generalmente ubicado en la parte frontal y de carácter direccional. Los valores se miden en lux, unidad del Sistema Internacional que cuantifica el flujo luminoso por unidad de área. Mediante una calibración experimental, se definieron seis rangos interpretativos: oscuro (≤ 4 lux), muy baja, baja, media, alta y muy alta (≥ 1000 lux). El sistema compara la lectura del sensor y anuncia la categoría correspondiente.

3.3.8 Detección y reconocimiento de marcadores QR

Una opción para la detección y decodificación de marcadores da al usuario la posibilidad de descubrirlos, enfocarlos, conocer y usar su contenido. De manera similar a la navegación por etiquetas, se indica la posición del marcador, no así su distancia. Una vez decodificado, su contenido se dice mediante la voz sintetizada. Si es una web se da al usuario la posibilidad de abrirla.

3.3.9 Ampliación de imagen

Finalmente, para usuarios con baja visión, se incluye una función de ampliación de imagen que amplía al doble de su tamaño la vista capturada por la cámara. La imagen ampliada puede visualizarse en cinco modos de representación: original, inversa, alto contraste, alto contraste inverso y amarillo sobre negro, este último especialmente diseñado para maximizar la legibilidad bajo condiciones visuales limitadas.

3.4 Interactividad y Realimentación

El diseño de interacción de la aplicación se fundamenta en un enfoque centrado en el usuario con discapacidad visual, priorizando sus necesidades específicas para garantizar accesibilidad, usabilidad y eficiencia. Este enfoque se validó mediante pruebas con usuarios, orientadas a refinar la retroalimentación, los métodos de entrada y la lógica de navegación [32, 33].

La interacción principal se basa en gestos táctiles intuitivos, alineados con los patrones ya familiarizados por los usuarios en otras aplicaciones de asistencia. Mediante deslizamientos horizontales, el usuario navega entre las funciones de reconocimiento; con deslizamientos verticales, accede a secciones de ayuda e información. Cada sección se identifica mediante síntesis de voz (TTS), anunciando su propósito, opciones disponibles y contexto actual, lo que permite una orientación clara sin dependencia visual. La navegación sigue una estructura cíclica, facilitando la exploración.

La activación de funciones se realiza mediante un toque prolongado en la pantalla, confirmado inmediatamente por una señal háptica (vibración) y una secuencia de tonos cortos espaciados, que persisten mientras la función permanece activa, manteniendo al usuario informado del estado del sistema.

Complementariamente, la aplicación integra un sistema de reconocimiento de comandos de voz basado en el *toolkit* VOSK, que permite operarla sin interacción táctil. Este modo permanece activo incluso cuando la aplicación está en segundo plano (opcional) o la pantalla está apagada, facilitando su uso en movimiento o con las manos ocupadas. Los comandos constan de una palabra de activación (variantes fonéticas de “asistencia”: asiste, asistir, asista, etc.) seguida de una o más palabras de mando. Tras detectar la palabra de activación, el sistema emite tonos de escucha activa y espera brevemente una instrucción válida (p. ej., “reconocer pesos”, “activar pesos”, “hora”, “minimizar”). Si no se recibe un comando reconocible en el intervalo definido, se notifica el fin de la escucha, requiriendo una nueva activación.

La retroalimentación multimodal es un pilar del diseño: se entrega principalmente mediante voz sintetizada a través del motor TTS del sistema, disponible desde Android 1.6, pero también mediante señales sonoras codificadas (tonos de inicio, fin, estado continuo) y retroalimentación háptica. Para usuarios con baja visión, se incluye opcionalmente retroalimentación con indicadores en alto contraste representados en la pantalla. Con el fin de evitar solapamientos en dispositivos de alto rendimiento, se introduce una pausa controlada entre respuestas sucesivas, asegurando claridad en la comunicación sonora.

Adicionalmente, se recomienda el uso de audífonos transductores óseos en entornos ruidosos o durante la navegación exterior, ya que transmiten el audio mediante vibración sin ocluir el canal auditivo, preservando la percepción de sonidos ambientales esenciales para la orientación.

Toda la información contextual y de ayuda se vocaliza mediante el TTS del dispositivo, mientras que la documentación completa se mantiene disponible en el manual de usuario y en la página web oficial, asegurando accesibilidad integral sin sobrecargar la interfaz en tiempo de uso.

3.5 Privacidad

En un contexto donde la seguridad y la privacidad han adquirido una relevancia crítica en el desarrollo de aplicaciones informáticas, el procesamiento de datos en servidores externos, con los riesgos asociados de almacenamiento indefinido, falta de control del usuario y exposición a terceros, representa una preocupación significativa.

Ante este escenario, la aplicación propuesta adopta un enfoque estrictamente local: no realiza ninguna conexión a servicios externos, ni transmite, almacena ni procesa datos fuera del dispositivo del usuario, y todas las operaciones, incluyendo la captura, análisis y descarte inmediato de imágenes, se ejecutan íntegramente en el terminal, sin persistencia de datos.

Esta arquitectura garantiza máxima privacidad y autonomía operativa, especialmente valiosa en entornos con conectividad limitada o costosa. No obstante, implica desafíos técnicos: la necesidad de integrar modelos de inteligencia artificial, bibliotecas de visión por computadora y recursos lingüísticos directamente en la aplicación incrementa su tamaño de instalación y demanda mayor consumo energético durante la ejecución de funcionalidades intensivas en cómputo, como el reconocimiento visual o el procesamiento de voz. Estos costos se asumen deliberadamente como un compromiso necesario para preservar la confidencialidad y la independencia del usuario con discapacidad visual.

4. -Resultados

En secciones anteriores se han presentado algunos resultados cuantitativos de las funciones implementadas. Los resultados globales y calificación de la solución no han sido aún cuantificados. Sin embargo, la solución anterior (qBillete), mucho más limitada y menos pulida, obtuvo una calificación de 4.5 puntos de 5 (90% de aprobación general) en una encuesta respondida por más de 40 de sus usuarios.

Las pruebas con usuarios constituyen una etapa fundamental en el desarrollo de aplicaciones orientadas a personas con discapacidad visual, permitiendo evaluar la interactividad, funcionalidad, reactividad del sistema y la eficacia de la retroalimentación, además de identificar errores y oportunidades de mejora en accesibilidad y usabilidad. En este caso, se realizaron pruebas con un grupo diverso de voluntarios, ciegos, con baja visión y videntes, incluyendo usuarios experimentados en el manejo de tecnologías asistivas. Todos contaban con lectores de pantalla activos en sus dispositivos, lo que permitió verificar la compatibilidad de los mecanismos de interacción táctil y de síntesis de voz.

Se evaluó también la compatibilidad con una amplia gama de dispositivos Android, con modelos tan antiguos como algunos lanzados en 2014, y con distintos niveles de rendimiento computacional. Este enfoque responde a la realidad socioeconómica del grupo poblacional objetivo, que frecuentemente utiliza hardware de gama baja o desfasado. Los participantes reportaron detalladamente los incidentes, especificando las operaciones realizadas, el software auxiliar empleado y las características técnicas de sus terminales, además de recomendaciones para adaptar mejor la aplicación a sus hábitos de uso.

Los problemas más reportados han estado relacionados con fallos en ejecución, usualmente dados por la gran diversidad de versiones del sistema operativo Android. Otro problema encontrado fueron algunas incompatibilidades con los gestos de sistemas interactivos ya instalados en los dispositivos.

Las funcionalidades menos robustas, y en las que hay que trabajar más, se determinaron como las de reconocimiento de textos y de colores.

La aplicación se distribuye sin restricciones mediante una página web dedicada, un canal de Telegram con chat integrado para retroalimentación, y la plataforma nacional de aplicaciones móviles Apklis. No requiere registro, activación ni validación remota, y su instalación permite la redistribución libre entre usuarios. Esta estrategia ha facilitado una amplia adopción, corroborada mediante intercambios con representantes de la Asociación Nacional del Ciego (ANCI), quienes confirman una

difusión significativa de esta y versiones anteriores en todo el país. Los comentarios recibidos reflejan una alta aceptación, destacando el impacto positivo en la autonomía, seguridad e inclusión social de los usuarios.

La aplicación resultante cumple con las pautas marcadas durante su concepción. Desde su introducción en 2022 ha sido actualizada una decena de veces, realizando mejoras y correcciones, varias de ellas reportadas o sugeridas por sus usuarios.

5. -Conclusiones

La aplicación aSISTA ha sido diseñada específicamente para personas con discapacidad visual, con el objetivo de potenciar su independencia funcional, facilitar su desenvolvimiento cotidiano y promover su participación social. Los resultados de las pruebas y la retroalimentación continua evidencian una buena efectividad en tareas de reconocimiento, siempre que se sigan las pautas de uso, así como una interacción adaptada a las necesidades sensoriales y cognitivas de este grupo. La compatibilidad está garantizada en la mayoría de los dispositivos Android lanzados en los últimos ocho años, lo que asegura accesibilidad en contextos de recursos limitados. El plan de desarrollo futuro contempla:

- reconocimiento más robusto de texto, y en documentos estructurados,
- descripción semántica más rica del entorno,
- un detector de marcadores fiduciarios más robusto, empleando el modelo Yunet [34],
- robustecer la estimación de distancias con apoyo adicional de un modelo de estimación mono-ocular,
- alertas de colisión durante la marcha mediante el sensor de aceleración,
- navegación asistida mediante GPS y datos geográficos,
- detector de marcadores QR más robusto,
- reconocimiento de color más robusto, basado en una red neuronal,
- modelos para reconocimiento y descripción de personas,
- expansión del vocabulario de comandos de voz,
- potencial extensión a otras áreas de asistencia, incluyendo el reconocimiento de lenguajes de señas.

En conjunto, aSISTA se erige como una plataforma inteligente integral, que empodera a las personas con discapacidad visual al otorgarles la capacidad de percibir, interpretar e interactuar con su entorno de formas que antes eran extremadamente difíciles o inaccesibles, contribuyendo significativamente a su autonomía y calidad de vida.

REFERENCIAS

1. World Health Organization. World report on vision. Exec. Summary. Department of Non communicable Diseases, 2019
2. Rodríguez, A., et al. Open Challenges of Blind People Using Smartphones. *International Journal of Human-Computer Interaction*, 36:17, 1605-1622, 2020
3. W3C. Web Content Accessibility Guidelines (WCAG) 2.1. W3C Recommendation 05 June 2018
4. Augustin, A. et al. Anxiety and depression prevalence rates in age-related macular degeneration. *Investigative Ophthalmology and Vision Science*, 48, 1498-1503, 2007
5. 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
6. Khan, A., Khusro, S. An insight into smartphone-based assistive solutions for visually impaired and blind people: issues, challenges and opportunities. *Universal Access in the Information Society* 20, pp. 265-298, 2021
7. Bhagat S, Joshi P, Agarwal A, Gupta S. Accessibility evaluation of major assistive mobile applications available for the visually impaired. *arXiv:2407.17496v1 [cs.HC]*, 2023
8. Al-Razgan, M., Almoaiqel, S., Alrajhi, N., et al. A systematic literature review on the usability of mobile applications for visually impaired users. *PeerJ Comput Sci.* 2021;7:e771
9. Gupta, S., Bhagat, S., Joshi, P., Agarwal, A. Accessibility evaluation of major assistive mobile applications. *ITU Journal on Future and Evolving Technologies*, Volume 4, Issue 4, December 2023
10. Cruz, F., et al. Smart backpack for the blind with light sensors, zigbee, rfid for grid-based election. *AIP Conference Proceedings*. vol. 2045, p. 20-54. AIP Publishing, 2018
11. Sato, D., Oh, U., Naito, K., Takagi, H., Kitani, K., Asakawa, C. Navcog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment. *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. pp. 270-279. ACM, 2018

12. Hossain, E., Rahman, M., Qaiduzzaman, K. Sightless Helper: An Interactive Mobile Application for Blind Assistance and Safe Navigation. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, volume 325, Cyber Security and Computer Science (pp.581-592), 2020
13. Rashid, A., Prati, A., Cucchiara, R. On the design of embedded solutions to banknote recognition. Optical Engineering 52(9), pp. 93-106, 2013
14. Solymar, Z., Stubendek, A., Radvanyi, M., Karacs, K. Banknote recognition for visually impaired. 20th European Conference on Circuit Theory and Design, 2011
15. Dobosz, K. Designing Mobile Applications for Visually Impaired People. Visually Impaired: Assistive Technologies, Challenges and Coping Strategies (pp.103-126), Nova Science Pub, 2017
16. Salas, M. Asistencia a discapacitados visuales para reconocer billetes mediante aplicación móvil. Revista Cubana de Ciencias Informáticas, Vol. 15, No. Especial UCIENCIA II, IX Taller Internacional de Software Libre, septiembre, 2021,
17. Salas, M.. qBillete. Asistencia en dispositivos móviles a personas con discapacidades visuales. Memorias de la Convención y Feria Informática 2022. InfoTIC
18. Leo, M., Medioni, G., Trivedi, M., Kanade, T., Farinella, G.M. Computer vision for assistive technologies. Computer Vision and Image Understanding 154, 1–15, 2017
19. 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
20. 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. HCI 2020, Springer
21. Howard, A., Sandler, M. et al. Searching for MobileNetV3. 2019 IEEE/CVF International Conf. on Computer Vision
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, 2017
23. Wang, C.Y., Liao, H.Y. M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. arXiv preprint arXiv:2402.13616, 2024
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, 2001
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. Zhan, H., Lyu, S., Lu, Y., Pal, U.. DenseNet-CTC: An end-to-end RNN-free architecture for context-free string recognition. Computer Vision and Image Understanding, Volume 204, March 2021
30. Ooms, J.. tesseract: Open Source OCR Engine. Disponible en : <https://github.com/ropensci/tesseract>, 2025
31. 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. VOSK API, Alphacep, 2025
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
33. Rogers, J., Sharp, Y., Preece, H. Interaction design: beyond human computer interaction. 6th Edition, Wiley, 2023.
34. Wu, W., Peng, H., Yu, S. Yunet: A tiny millisecond-level face detector. Machine Intelligence Research 20, 2023

CONFLICTO DE INTERESES

Ninguno de los autores manifestó la existencia de posibles conflictos de intereses que debieran ser declarados en relación con este artículo.

AUTORES

Maikel Salas Zaldivar, Ingeniero en Informática, La Habana, Cuba, e-mail: maikelsz8143@nauta.cu

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



Esta revista se publica bajo una [Licencia Creative Commons Atribución-No Comercial-Sin Derivar 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/)