



UNIVERSIDAD TECNOLÓGICA ISRAEL

TRABAJO DE TITULACIÓN EN OPCIÓN AL GRADO DE:

INGENIERO EN ELECTRÓNICA DIGITAL Y TELECOMUNICACIONES

TEMA: DESARROLLO DE UN DISPOSITIVO QUE MEDIANTE VISIÓN ARTIFICIAL PERMITA ADQUIRIR IMÁGENES CON PALABRAS PARA LA CONVERSIÓN A AUDIO, ORIENTADO A LA AYUDA DE PERSONAS INVIDENTES.

AUTOR: JORGE LUIS CHINCHERO IZA

TUTOR: ING. FIDEL DAVID PARRA BALZA, PhD

QUITO- ECUADOR

AÑO: 2019

UNIVERSIDAD TECNOLÓGICA ISRAEL

APROBACIÓN DEL TUTOR

En mi calidad de tutor del trabajo de titulación certifico:

Que el trabajo de titulación **DESARROLLO DE UN DISPOSITIVO QUE MEDIANTE VISIÓN ARTIFICIAL PERMITA ADQUIRIR IMÁGENES CON PALABRAS PARA LA CONVERSIÓN A AUDIO, ORIENTADO A LA AYUDA DE PERSONAS INVIDENTES.**, presentado por el **Sr. Jorge Luis Chinchero Iza**, estudiante de la carrera de **Electrónica Digital y Telecomunicaciones**, reúne los requisitos y méritos suficientes para ser sometido a la evaluación del Tribunal de Grado, que se designe, para su correspondiente estudio y calificación.

Quito D.M. junio de 2019

TUTOR

Ing. Fidel David Parra Balza, PhD.

AGRADECIMIENTO

Agradezco a Dios por la vida y por brindarme la sabiduría necesaria para alcanzar esta meta tan importante en el ámbito de mi vida profesional.

A mi primo Cristian por haberme apoyado en todos los momentos que he necesitado, por ofrecerme una palabra de aliento en los momentos críticos, gracias por tu amistad.

A mi familia que se han preocupado por mi bienestar y el de mi madre, gracias por todo su apoyo e interés.

Finalmente, a mis amigos, con quienes compartí buenos y malos momentos en la universidad, gracias por seguir contando con su valiosa amistad.

De todo corazón gracias por el apoyo que recibí de cada uno de ustedes.

Jorge Luis

DEDICATORIA

A mi madre Margoth, que, a pesar de mis errores y fracasos, me apoyo en cada decisión que tome, sin importar cuál sea. Gracias por toda la paciencia, por los años de sacrificio, la dedicación y por el inmenso amor que me transmitiste al momento de enseñarme cosas nuevas. ¡Te amo mamita!

A mi padre José, gracias por cuidarme desde el cielo, y por tus sabias palabras que siempre las tengo presente a cada instante de mi vida.

A mi novia Johanna, gracias por toda la paciencia, comprensión y amor a lo largo de este camino, por ser mí apoyo incondicional y mi personita especial.

A ustedes dedico cada esfuerzo, ya que, gracias a su ayuda, paciencia y amor, logre culminar una etapa más de mi vida. Sin ustedes no soy nada.

Jorge Luis.

ÍNDICE DE CONTENIDOS

APROBACIÓN DEL TUTOR	I
AGRADECIMIENTO	II
DEDICATORIA	III
ÍNDICE DE CONTENIDOS	IV
ÍNDICE DE FIGURAS	VIII
ÍNDICE DE TABLAS	X
ÍNDICE DE ECUACIONES	XI
RESUMEN	XII
ABSTRACT	XIII
INTRODUCCIÓN	1
Antecedentes de la situación objeto de estudio	1
Planteamiento y justificación del problema.....	3
Objetivo General.....	3
Objetivo Específico	4
Beneficiarios	4
Alcance	4
Descripción de los capítulos	5
CAPÍTULO 1	6

FUNDAMENTACIÓN TEÓRICA	6
1.1 Visión artificial.....	6
1.2 Procesamiento digital de imágenes.....	7
1.2.1 Conceptos básicos del procesamiento de imágenes.	8
1.2.2 Reconocimiento de patrones.....	9
1.3 Iluminación.....	9
1.3.1 Iluminación frontal.....	10
1.3.2. Iluminación lateral.....	11
1.4 Raspberry Pi.....	12
1.4.1 Pines GPIO	14
1.5 Lenguaje de programación Python.....	15
1.6 Librería OpenCV.....	15
1.6.1.- Estructura.....	16
1.7 Conversor de texto a voz.....	17
1.8 Equipo Reconocimiento óptico de caracteres (OCR).....	17
1.9 Cámara web.....	17
1.10.- Librería Espeak texto a voz.....	18
1.11.- Librería Python – Tesseract.....	18
CAPÍTULO 2	19
MARCO METODOLÓGICO	19
2.1 Tipo y diseño de investigación.....	19
2.2 Técnicas e instrumentos de recolección de datos	20
2.3 Metodología seleccionada.....	21
CAPÍTULO 3	24

PROPUESTA	24
3.1 Misión.....	24
3.2 Justificación	25
3.3 Factibilidad de la propuesta.....	25
3.3.1 Estudio técnico.	25
3.3.2 Características de los componentes.	36
3.3.3 Parámetros de la cámara.....	39
3.4 Estudio económico.	42
3.5 Conclusión de la propuesta.....	45
CAPÍTULO 4	46
IMPLEMENTACIÓN	46
4.1 Desarrollo.	46
4.1.1 Proceso de instalación del sistema operativo Raspbian en la tarjeta microSD.	46
4.1.2 Etapa de acondicionamiento y almacenamiento de la imagen.	49
4.1.3 Etapa de diseño del algoritmo.	50
4.1.4 Procesos realizados por el algoritmo.....	50
4.1.4.1 Etapa de procesamiento de la imagen.....	52
4.1.4.2 Etapa de reconocimiento óptico de caracteres (OCR).....	54
4.1.4.3 Etapa de conversión a audio.	55
4.2 Implementación.	56
4.3 Pruebas de funcionamiento.....	58
4.3.1 Ambiente de prueba.....	58
4.3.2 Pruebas del prototipo.....	59
4.4 Análisis de resultados.	61

CONCLUSIONES.....	62
RECOMENDACIONES	64
REFERENCIAS BIBLIOGRÁFICAS	65
ANEXOS.....	67

ÍNDICE DE FIGURAS

Figura 1.1. Elementos de un sistema de visión artificial.....	7
Figura 1.2. Espectro visible para las cámaras.....	9
Figura 1.3. Iluminación frontal.....	11
Figura 1.4. Iluminación lateral.	11
Figura 1.5. Raspberry Pi 3 B.	12
Figura 1.6. Estructura Raspberry Pi.....	14
Figura 1.7. Distribución de pines GPIO.	14
Figura 1.8. Estructura de la librería OpenCV.....	16
Figura 1.9. WebCam C922 PRO HD STREAM.	18
Figura 3.1. Umbral fijo, función thresh.....	29
Figura 3.2. Umbral binario, función thresh.	29
Figura. 3.3. Circuito para la captura de imágenes.	38
Figura 3.4. Campo de visión y distancia de trabajo.....	40
Figura 4.1. Gestor de instalación de sistemas operativos Noobs.	47
Figura 4.2. Cuadro de dialogo SD Card Formatter.....	47
Figura 4.3. Elección del sistema operativo Raspbian.....	48
Figura 4.4. Proceso de instalación del sistema operativo Raspbian.	48
Figura 4.5. Finalización de la instalación del sistema operativo Raspbian.	49

Figura 4.6. Pantalla de inicio del sistema operativo Raspbian.	49
Figura 4.7. Circuito para la captura de imágenes – pulsador.	50
Figura 4.8. Diagrama de los procesos realizados por el algoritmo.....	51
Figura 4.9. Diagrama del procesamiento de la imagen.	53
Figura 4.10. Diagrama del reconocimiento óptico de caracteres.	55
Figura 4.11. Diagrama de la reproducción del texto en voz sintetizada.....	56
Figura 4.12. Fuente de alimentación.	57
Figura 4.13. Carcasa del sistema conectado.	57
Figura 4.14. Imagen del sistema convertidor de palabras a audio.....	58
Figura 4.15 Cuadro de dialogo del algoritmo implementado.....	59
Figura 4.16. Captura de la región de interés de la imagen.	60
Figura 4.17. Conversión a escala de grises.....	60
Figura 4.18. Extracción de caracteres.....	60

ÍNDICE DE TABLAS

Tabla 3.1. Cuadro comparativo de las cámaras web consideradas.....	26
Tabla 3.2. Características técnicas de la cámara web.....	27
Tabla 3.3. Software disponible para el reconocimiento óptico de caracteres.....	27
Tabla 3.4. Descripción del comando espeak	31
Tabla 3.5. Primera encuesta.....	32
Tabla 3.6. Segunda encuesta.....	34
Tabla 3.7. Tercera encuesta.	35
Tabla 3.8. Recursos tecnológicos (SOFTWARE)	43
Tabla 3.9. Recursos tecnológicos (COMPONENTES ELECTRÓNICOS)	43
Tabla 3.10. Componentes adicionales.	44
Tabla 3.11. Mano de obra	44
Tabla 3.12. Valor total del proyecto.	44
Tabla 4.1. Documentos analizados	61

ÍNDICE DE ECUACIONES

Ecuación 1. Modelamiento matemático de la función thresh, binario.	29
Ecuación 2. Ecuación para el cálculo de la capacidad de una batería.	37
Ecuación 3. Ecuación para el cálculo del valor de la resistencia pull - down máxima.	38
Ecuación 4. Ecuación para el cálculo del valor de la resistencia pull-up mínima.	38
Ecuación 5. Ecuación para el cálculo del área de un rectángulo.	39
Ecuación 6. Ecuación del campo visual (FOV).	40

RESUMEN

El presente documento consiste en el desarrollo de un dispositivo que mediante visión artificial permita convertir las palabras contenidas en una imagen en audio, orientado a la ayuda de personas invidentes, presenta inicialmente los conceptos teóricos sobre la **visión artificial**, captura de imágenes, **procesamiento de imágenes**, **reconocimiento óptico de caracteres (OCR)** y el **lenguaje de programación Phyton**.

Para el desarrollo del proyecto se utilizó una tarjeta **Raspberry Pi**, en la cual se realizó el código de programación de los diferentes procesos necesarios para la conversión de imágenes con palabras a audio. De igual manera se empleará otros elementos para el suministro de voltaje, captura de imágenes y reproducción de la voz sintetizada en auriculares.

Así también, se encuentran los datos obtenidos al realizar las pruebas de funcionamiento juntamente con el análisis de los mismos, las mismas serán realizadas con una persona invidente.

Finalmente, la conclusión más relevante de dispositivo cumple con el objetivo de ser portátil, además de considerarse que en base a las pruebas realizadas con el dispositivo se logró la conversión a audio de las imágenes, y dicho dispositivo ayudó con la lectura de palabras.

Palabras clave: visión artificial, raspberry pi, procesamiento de imágenes, phyton, reconocimiento óptico de caracteres.

ABSTRACT

The present document "Development of a device that by means of artificial vision allows to convert the words contained in an audio image, oriented to the help of blind people", initially presents the theoretical concepts on **artificial vision**, image capture, **image processing**, **Optical Character Recognition (OCR)** and the **Phyton programming language**.

For the development of the project a **Raspberry Pi** card will be used, in which the programming code of the different processes necessary for the conversion of images with words to audio was made. In the same way, other elements will be used to supply voltage, capture images and reproduce the synthesized voice in headphones.

Also, the data obtained by performing the functional tests together with the analysis of the same, will be performed with a blind person.

Finally, the most relevant conclusion of the device fulfills the objective of being portable, besides considering that based on the tests carried out with the device, the conversion to audio of the images was achieved, and said device helped with the reading of words

Keywords: artificial vision, raspberry pi, image processing, phyton, optical character recognition.

INTRODUCCIÓN

Antecedentes de la situación objeto de estudio

La mayoría de personas invidentes tradicionalmente utiliza un bastón para desplazarse, otras utilizan un perro guía. Sin embargo, estos aditamentos tienen sus limitaciones y desventajas. Los dispositivos electrónicos de apoyo a la movilidad han tenido poco éxito comercialmente ya que deberían competir con la sencillez y bajo costo del bastón, dando como resultado que estos dispositivos no sean comercializados.

En los últimos años se ha desarrollado tecnologías para la asistencia de personas invidentes, algunos usan sensores ultrasónicos, infrarrojos o laser, orientado a la detección de objetos. La visión artificial, a diferencia de estas tecnologías permite una interpretación cognitiva del entorno, ofreciendo un mayor grado de reproducción de la realidad.

El desarrollo de técnicas de reconocimiento de caracteres (OCR), ha permitido el surgimiento de dispositivos para acceder a libros, revistas, señales en la calle e información en general. En un inicio estos dispositivos eran grandes y necesitaban computadoras para escanear toda la hoja. Actualmente se encuentran estos principios en teléfonos inteligentes. El problema de estas aplicaciones es que resulta complicado para el usuario invidente apuntar la cámara para encuadrar correctamente el texto. Actualmente, las investigaciones en este campo están enfocadas en detectar texto en imágenes como son: señales en la calle, frases, entre otras.

En Ecuador la investigación de la visión artificial orientada específicamente a la ayuda de personas con discapacidad visual es muy escasa, debido a su alto costo y su difícil manipulación por parte de la persona invidente. A continuación, se cita algunos estudios en el ámbito de desarrollo de prototipos para la ayuda de personas invidentes:

(Ayala & Vásquez, 2012), de la Universidad Politécnica Salesiana (UPS), desarrollaron un prototipo de un sistema de ultrasonido aplicado a un bastón blanco, mismo que permite a personas invidentes detectar obstáculos mediante un sistema de ultrasonido que comunica, a través de audífonos, los objetos que la persona tiene por delante. Esta investigación contribuyó al presente proyecto, en el ámbito de la detección de objetos y las técnicas empleadas para la reproducción de audio.

(Ñiacasha, 2004), de La Escuela Politécnica Nacional, en su proyecto de tesis de Ingeniería en Electrónica y Control titulado “Desarrollo de un dispositivo que mida la distancia a un objeto emulando el efecto de un bastón blanco para personas invidentes”, mismo que proporciona la información que brinda un bastón sin la necesidad del mismo y envía información mediante vibraciones de diferentes frecuencias hacia la mano del usuario. Este proyecto sirvió de ayuda para el análisis de las diferentes ayudas existentes para la movilización de las personas invidentes, además del análisis de los principios empleados por los sensores de distancia existentes en el mercado.

Con ayuda del sistema braille las personas invidentes pueden acceder vía tacto a dicha información, que en algunos casos se vuelve dificultoso aprender nuevos sistemas para personas de edades avanzadas. Existen diferentes equipos adaptados que transforman palabras a audio, así también la existencia de textos electrónicos, pero son limitados y requieren equipos adicionales para su funcionamiento.

Sin embargo, en el ámbito de diseño de prototipos que faciliten la movilidad de personas invidentes es costosa y escasa, existiendo en el mercado algunas opciones que ayudan a detección de objetos, rostros, obstáculos e incluso poder conocer el color y formas de algunas cosas que para las personas que gozan de la capacidad de ver, resulta algo cotidiano.

Basado en las investigaciones anteriores y tomando en cuenta la forma en la que aborda el problema se presenta la actual investigación con la finalidad de facilitar el acceso a la información, específicamente de textos cortos que brindan mayor información, en comparación de los textos en braille que no poseen algunas combinaciones, signos y letras específicas del idioma, como la ñ del español.

Planteamiento y justificación del problema

En Ecuador, según las cifras del Consejo Nacional de Igualdad de Discapacidades, asegura que, para el presente mes de noviembre de 2018, existen 53611 personas con discapacidad visual, de los cuales 8385 radican en la provincia de Pichincha. Una persona con discapacidad visual es aquella que presenta una ausencia total de percepción visual o percibir luz sin lograr definir que es o de donde proviene. Estas carencias visuales influyen directamente en los aspectos físicos, emocionales, sociales y profesionales. Dichos impedimentos dificultan a las personas a adquirir nuevos aprendizajes y a realizar las actividades cotidianas como es: la comunicación, orientación y movilidad. Debido a estos impedimentos las personas con discapacidad visual se ayudan utilizando la información que obtienen de sus otros sentidos.

La existencia de software que convierte palabras en audio requieren de un computador para la instalación de dicho software y posterior conversión, además que requiere de algunas operaciones que una persona invidente no puede realizar con facilidad, las aplicaciones para teléfonos inteligentes se vuelven dificultosas ya que requiere necesariamente la capacidad de ver para activar la aplicación. Al no poseer un dispositivo de visión artificial que ayude a las personas con discapacidad visual a acceder a la información específicamente a la lectura, ya que el llamado bastón, perro lazarillo y otros dispositivos utilizados para este fin no pueden detectar palabras para conversión a audio.

Para lo cual se plantea el uso de la técnica de visión artificial para la adquisición y procesamiento de imágenes con palabras y posterior conversión a audio sobre la tarjeta Raspberry Pi, lo que permitirá disminuir costes de diseño y servirá como ayuda en el campo de la interpretación de texto para las personas invidentes.

Objetivo General

- Desarrollar un dispositivo que mediante visión artificial permita adquirir imágenes con palabras para la conversión a audio, orientado a la ayuda de personas invidentes.

Objetivo Específico

- Establecer parámetros de funcionamiento de la tarjeta de desarrollo Raspberry Pi para la adquisición de imágenes y posterior procesamiento.
- Adquirir de una imagen de tamaño 21 cm x 29 cm aproximadamente, palabras tipo Times New Roman, número 18, en español mediante la utilización de una cámara web, para su posterior digitalización.
- Procesar las palabras adquiridas mediante la cámara web para la conversión a audio.
- Diseñar e Implementar la estructura y parte electrónica sobre una gorra para el uso portátil del equipo.
- Realizar las pruebas de funcionamiento del equipo.

Beneficiarios

BENEFICIARIOS DIRECTOS: Al ser un dispositivo conversor de texto a audio, beneficia directamente a la persona que está ejecutándolo al tratarse de un proyecto de titulación. A la vez se beneficiarán los proveedores nacionales e internacionales de los diferentes elementos que componen el proyecto.

BENEFICIARIOS INDIRECTOS: El propósito que atrae el desarrollo de la propuesta enfocadas al diseño de un dispositivo conversor de texto a audio, beneficia indirectamente a quienes va dirigido el proyecto en este caso particular a las personas invidentes. También será de utilidad para futuras investigaciones para personas que quieran realizar tecnologías más simples para el tratamiento de imágenes con palabras y el reconocimiento óptico de caracteres.

Alcance

El sistema a implementar está enfocado en el tratamiento de imágenes con un tamaño de 21 cm x 29 cm aproximadamente, con palabras tipo Times New Roman, número 18, en español mediante la utilización de una cámara web. Este sistema se configurará para que las imágenes sean captadas únicamente en el día o cuando exista una adecuada iluminación, las palabras contenidas en la imagen se procesaran en un tiempo aproximado de 1 minuto y se presentaran a través de un auricular para ser escuchadas por el usuario.

Descripción de los capítulos

El primer capítulo describe la problemática, justificación, objetivos y la limitación del proyecto y de cómo la implementación de este sistema servirá de ayuda a la movilidad de personas invidentes.

El segundo capítulo describe la metodología que se aplicara para la elaboración del proyecto. Además de las características, ventajas, desventajas, funcionamiento del sistema y los elementos necesarios para el procesamiento de imágenes. También se hace mención el uso de la plataforma tanto de hardware libre llamado Raspberry Pi y a la vez el software llamado Phyton; la teoría y características del código de línea para su aplicación en sistemas de visión artificial.

El tercer capítulo hace referencia al tratamiento de las imágenes y de las diferentes etapas que la imagen debe atravesar, como son: etapa sensorial, etapa de procesado, etapa de segmentación, etapa de parametrización, etapa de clasificación, y etapa de interpretación; propios de la visión artificial y que servirán para el propósito de este proyecto. Finalmente se presentará el equipo ensamblado sobre una gorra que constará de la tarjeta Rspberry Pi, la cámara web, y el auricular.

El cuarto capítulo, se presenta la implementación y las pruebas realizadas en ambientes reales que permitirán recolectar información, para posibles calibraciones del prototipo tanto en el aspecto de la adquisición de imágenes, así como la obtención del audio en español.

CAPÍTULO 1

FUNDAMENTACIÓN TEÓRICA

El presente capítulo aborda los conceptos fundamentales sobre una de las técnicas del procesamiento de imágenes que permitirán la ejecución e implementación del presente proyecto, así también, se hará una explicación de las diferentes características. A la vez se describen los dispositivos y diferentes elementos empleados en el sistema a implementarse.

1.1 Visión artificial.

La visión artificial es un sub campo de la inteligencia artificial que, a través de las técnicas adecuadas, permite la obtención procesamiento y análisis de cualquier tipo de información especial obtenida a través de imágenes digitales.

El principio de funcionamiento de un sistema de visión artificial figura 1.1, consiste en la captación de imágenes a través de sensores (cámaras) para su posterior tratamiento a través de técnicas de procesamiento avanzadas, permitiendo así intervenir sobre un proceso o sistema. (Murillo Quishpe & Montaluisa Pilatasig, 2010).

Un sistema de visión artificial se compone de los siguientes elementos:

- **Fuente de luz:** Constituye un aspecto importante ya que debe proporcionar unas condiciones de iluminación uniforme e independiente del entorno, facilitando, además, si es posible, la extracción de los rasgos de interés para una determinada aplicación.
- **Sensor de imagen:** Son elementos sensibles a la luz que cambian su señal eléctrica con relación a la intensidad luminosa que percibe. Además de encargarse de la recolección de la información y características del objeto de estudio.

- **Tarjeta de captura o adquisición de la imagen:** Encargada de transferir la imagen de la cámara a la memoria de la tarjeta con el fin de que esta pueda realizar el procesamiento adecuado de la imagen.
- **Algoritmo de análisis de imágenes:** Consiste en aplicar las transformaciones necesarias y extracción de información contenida en las imágenes, con el fin de obtener los resultados para los que han sido diseñados.
- **Computadora o módulo de proceso:** Se trata del sistema encargado de recibir y analizar, mediante los algoritmos adecuados, las imágenes a la velocidad necesaria para poder interactuar con el entorno en tiempo real. También es la parte del sistema encargada de implementar la funciones y objetivos para los que se haya diseñado el sistema.
- **Sistema de respuesta en tiempo real:** Con la información extraída el sistema puede tomar decisiones que afecten al algoritmo con el fin realizar una tarea para la cual fue diseñado el sistema.

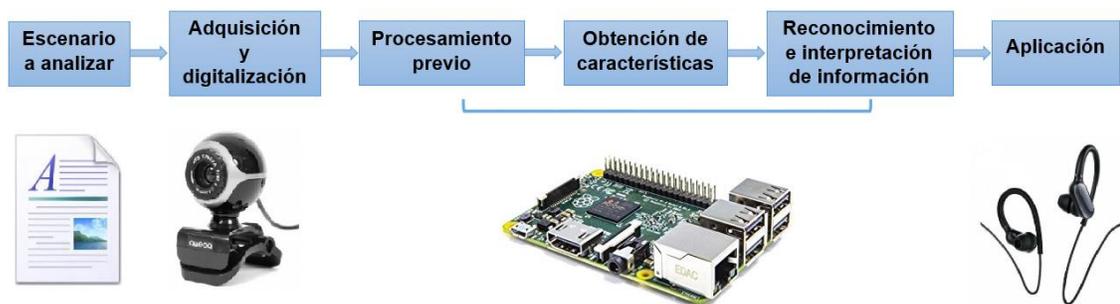


Figura 1.1. Elementos de un sistema de visión artificial.

Fuente. - Elaborado por el autor.

1.2 Procesamiento digital de imágenes.

Es el conjunto de técnicas que se aplican a las imágenes digitales con el objetivo de mejorar la calidad o facilitar la búsqueda de información. La imagen puede haber sido generada de muchas maneras, por ejemplo, fotográficamente, o por medio de monitores de televisión

El procesamiento digital de imágenes se ha trabajado desde dos enfoques. Uno desde el punto de vista de procesamiento digital de señales, en el cual se trabaja en el dominio de las frecuencias con ayuda de la transformada de Fourier. El otro en el dominio espacial donde se considera a la

imagen como una matriz, haciendo uso de algebra matricial. Ambos enfoques buscan producir una imagen de mejor calidad a partir de una imagen digital dada, se busca eliminar ruido o elementos que no permitan una imagen aceptable, para posteriormente extraer la información relevante.

1.2.1 Conceptos básicos del procesamiento de imágenes.

A continuación, se mencionan los conceptos básicos relacionados con el procesamiento de imágenes desde el enfoque de dominio espacial. (Esqueda Elizondo, 2002).

- **Visión por computadora:** Consiste en la adquisición, procesamiento, clasificación y reconocimiento de imágenes digitales.
- **Pixel:** Elemento básico de una imagen.
- **Imagen:** Arreglo bidimensional de píxeles con diferente intensidad luminosa (escala de grises). Si la intensidad luminosa de cada pixel se representa por n bits, entonces existirán 2^n escalas de gris diferentes.
- **Color:** Se forma mediante la combinación de los tres colores básicos: rojo, verde y azul (en inglés corresponde a las siglas RGB). Y puede expresarse mediante una tripleta de valores de 0 a 1, donde R, G y B representan las intensidades de cada uno de los tres colores básicos.
- **Brillo:** Indica si un área está más o menos iluminada.
- **Tono:** Indica si un área parece similar al rojo, amarillo, verde o azul, o una proporción de ellos.
- **Luminosidad:** Es el brillo de una zona respecto a otra zona blanca en la imagen.
- **Espacio RGB:** Se basa en la combinación de tres señales de luminancia cromática distinta: rojo, verde y azul (red, green, blue). La forma más sencilla de obtener un color específico es determinar la cantidad de color rojo, verde, azul que se requiere combinar para obtener el color deseado, para lo cual se realiza la suma aritmética de los componentes.
- **Histograma de la imagen:** Es una representación del número de píxeles de cierto nivel de gris en función de los niveles de gris.

1.2.2 Reconocimiento de patrones.

El reconocimiento de patrones también llamado lectura de patrones, identificación de figuras y reconocimiento de formas, es la ciencia que se ocupa de los procesos de ingeniería, computación y matemáticas relacionados con objetos físicos y/o abstractos con el propósito de extraer información que permita establecer propiedades de o entre conjuntos de dichos objetos.

1.3 Iluminación.

Las cámaras de color emulan la función del ojo humano utilizando una matriz compuesta de sensores que captan diferentes rangos del espectro visible. Normalmente se utilizan 3 rangos, el rojo (R), el verde (G) y el azul (B). Posteriormente, la información de cada uno de los grupos (RGB) que forma la matriz se agrupa en un solo pixel que contiene los tres valores Red-Green-Blue que ha obtenido el sensor. (DCM SISTEMES, 2018)

A continuación, la figura 1.2 muestra el espectro visible correspondiente a las cámaras.

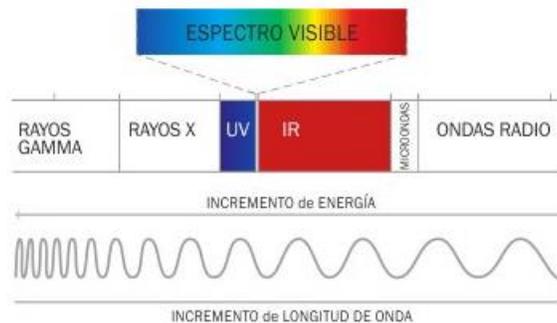


Figura 1.2. Espectro visible para las cámaras.

Fuente. - (DCM SISTEMES, 2018)

Dentro de un sistema de visión artificial la iluminación es un factor muy importante en la resolución de la aplicación. Las cámaras capturan la luz reflejada de los objetos para su posterior análisis.

El propósito de la iluminación es el de controlar la forma en que la cámara va a ver el objeto para determinar si cumple con las especificaciones requeridas.

Los objetivos de la iluminación son:

- Optimizar el contraste.
- Normalizar cualquier variación de la iluminación ambiente y simplificar el proceso de tratamiento posterior de la imagen (si se utilizan filtros por software el tiempo de procesado se ve incrementado).

Dentro del mercado se encuentran distintas opciones para la iluminación, a continuación, se citan algunos tipos:

- Halógenas.
- Incandescencia.
- Fluorescentes.
- Laser.
- Xenón.
- LED (Light Emitting Diode).

Existen distintas técnicas de iluminación para resaltar aspectos de los objetos a analizar. A continuación, se explica los sistemas de iluminación más utilizados en sistemas de visión artificial.

1.3.1 Iluminación frontal.

La cámara se posiciona mirando al objeto en la misma dirección que la luz. Esto reduce las sombras, suaviza las texturas y minimiza la influencia de rayas, polvo e imperfecciones que pueda tener el objeto. La cámara recibe la luz reflejada del objeto. Este tipo de iluminación se consigue mediante anillos de luz. (Nogue & Antiga, 2012).

La figura 1.3 hace referencia a la iluminación frontal, que tiene las siguientes características:

Aplicaciones: Indicada para superficies con pocos reflejos: papel, tela; para la detección de marcas de diferentes colores, caracteres y detección de todo lo que suponga un cambio de color en prácticamente cualquier superficie.

Ventajas: Elimina sombras, se puede utilizar a grandes distancias cámara/objeto.

Desventaja: Intensos reflejos sobre superficies reflectantes.

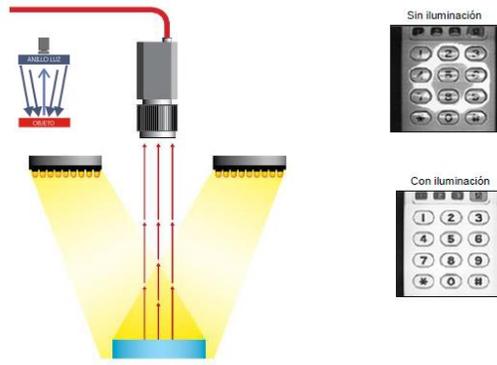


Figura 1.3. Iluminación frontal.

Fuente. - (Nogue & Antiga, 2012)

1.3.2. Iluminación lateral.

La cámara se posiciona mirando al objeto mientras que la dirección de la luz es lateral al objeto. El grado de inclinación del elemento emisor de luz vendrá determinado por el grado deseado de resalte de los relieves. (Nogue & Antiga, 2012)

La figura 1.4 muestra la iluminación lateral, que tiene las siguientes características:

Aplicaciones: indicada para resaltar bordes, rayas y fisuras en una dirección determinada.

Ventajas: resalta los relieves por pequeños que sean de los objetos, resultando una sombra muy definida.

Desventaja: con ángulos pequeños respecto a la horizontal, la luz producirá sombras en todos los relieves y en el contorno de la pieza.

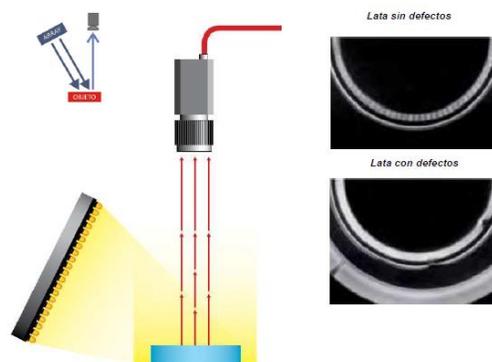


Figura 1.4. Iluminación lateral.

Fuente. - (Nogue & Antiga, 2012)

1.4 Raspberry Pi.

Para (Raspberry Shop, 2018) la Raspberry Pi es una computadora de placa reducida (Single Board Computer) de bajo costo desarrollada en Reino Unido por la Fundación Raspberry Pi, con el objetivo de estimular la enseñanza de ciencias de la computación en las escuelas. Se trata de una diminuta placa base de 85 x 54 milímetros en el que se aloja un chip Broadcom BCM2835 con procesador ARM hasta a 1 GHz de velocidad (modo Turbo haciendo overclock), GPU VideoCore IV y 512 Mbytes de memoria RAM. Las últimas placas como la Raspberry Pi 2 y Raspberry Pi 3 tienen 1GB de memoria RAM.

Para su funcionamiento, se necesita de un medio de almacenamiento (Raspberry Pi utiliza tarjetas de memoria SD o microSD dependiendo del modelo), conectarlo a la corriente utilizando cualquier cargador microUSB de al menos 1000 mah para las placas antiguas y de al menos 2500 mah para las modernas.

En cuanto a la conexión de red, se dispone de un puerto Ethernet (los modelos A y A+ no disponen de puerto Ethernet) para conectar un cable RJ-45 directamente al router o se puede recurrir a utilizar cualquier adaptador inalámbrico WiFi compatible. En este caso, la tarjeta empleada Raspberry Pi model B incorpora cuatro puertos USB, para poder conectar el teclado y el ratón.



Figura 1.5. Raspberry Pi 3 B.

Fuente. - (Raspberry Shop, 2018)

Como se muestra en la figura 1.5 la tarjeta Raspberry Pi 3 está construida alrededor del nuevo procesador BCM2837 ARMv8 de 64bits con 1,2GHz de velocidad, mucho más rápido y

con mayor capacidad de procesamiento que sus antecesores. Además, la nueva Raspberry Pi 3 integra el chip BCM43143 que la dota con conectividad Wifi b/g/n y Bluetooth 4.1 Low Energy de bajo consumo y cuenta con administración de energía mejorada que permite trabajar con más dispositivos USB externos.

La tarjeta Raspberry Pi 3 permite usar fuentes de hasta 2.5A para proveer más energía a los puertos USB. También se recomienda usar una fuente de alimentación de 5.1V y 2.5A para poder sacar todo el rendimiento.

Dentro de las principales especificaciones de la Raspberry Pi 3 model B se tiene:

- Chipset Broadcom BCM2837 a 1,2 GHz
- ARM Cortex-A53 de 64 bits y cuatro núcleos
- LAN inalámbrica 802.11 b/g/n
- Bluetooth 4.1 (Classic y Low Energy)
- Coprocesador multimedia de doble núcleo Videocore IV
- Memoria LPDDR2 de 1 GB
- Compatible con todas las últimas distribuciones de ARM GNU/Linux y Windows 10 IoT
- Conector micro USB para fuente de alimentación de 2,5 A
- 1 puerto Ethernet 10/100
- 1 conector de vídeo/audio HDMI
- 1 conector de vídeo/audio RCA
- 1 conector de cámara CSI
- 4 x puertos USB 2.0
- 40 pines GPIO
- Antena de chip
- Conector de pantalla DSI
- Ranura de tarjeta microSD
- Dimensiones: 85 x 56 x 17 mm

La figura 1.6 muestra la estructura de la tarjeta Raspberry Pi modelo 3 B

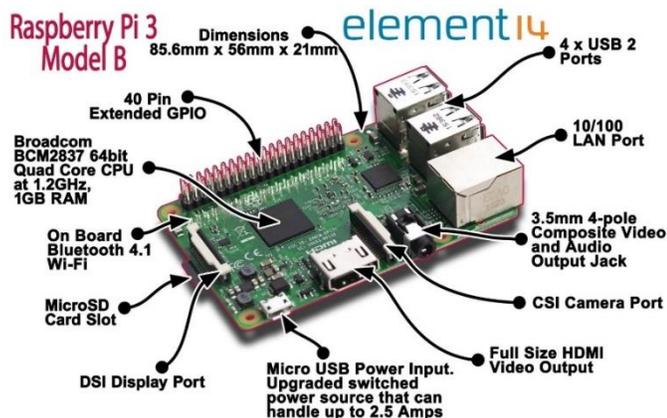


Figura 1.6. Estructura Raspberry Pi.

Fuente. - (Raspberry Shop, 2018)

1.4.1 Pines GPIO

Según (Raspberry Shop, 2018) el puerto GPIO (General Purpose Input/Output) es un sistema de E/S (Entrada / Salida) de propósito general, es decir, una serie de conexiones que se pueden usar como entradas y salidas. Los GPIO representan la interfz entre la tarjeta Raspberry Pi y el mundo exterior.

La distribución de pines GPIO varían según el modelo, en el presente caso se utilizó una tarjeta Raspberry Pi modelo 3 B, en la figura 1.7 se muestra la distribución de pines para el modelo antes señalado.

Raspberry Pi Model B
Rev 1 P1 GPIO Header

	Pin No.		
3.3V	1	2	5V
GPIO0	3	4	5V
GPIO1	5	6	GND
GPIO4	7	8	GPIO14
GND	9	10	GPIO15
GPIO17	11	12	GPIO18
GPIO21	13	14	GND
GPIO22	15	16	GPIO23
3.3V	17	18	GPIO24
GPIO10	19	20	GND
GPIO9	21	22	GPIO25
GPIO11	23	24	GPIO8
GND	25	26	GPIO7

Key

Power +	UART
GND	SPI
IC	GPIO

Figura 1.7. Distribución de pines GPIO.

Fuente. - (Raspberry Shop, 2018)

1.5 Lenguaje de programación Python.

Según (Palacios Sanchez, 2018) Python 2.7 está programado para ser la última versión principal de la serie 2.x antes de que pase a un periodo de mantenimiento extendido. Esta versión contiene muchas de las características que se lanzaron en Python 3.1.

Las mejoras en esta versión incluyen:

- Un tipo de diccionario ordenado.
- Un módulo io mucho más rápido.
- Mejoras Float Repr retroportadas desde 3.x
- Establecer y compresión de diccionarios.
- Vistas de diccionario.
- Nueva sintaxis para anidado con declaraciones.

Python cuenta con estructuras de datos eficientes y de alto nivel y un enfoque simple pero efectivo a la programación orientada a objetos. El lenguaje Python y sus librerías están a libre disposición en forma binaria y de código fuente para las principales plataformas desde el sitio web de Python, y puede distribuirse libremente.

Python puede extenderse fácilmente con nuevas funcionalidades y tipos de datos implementados en C o C++ (u otros lenguajes accesibles desde C). Python también puede usarse como un lenguaje de extensiones para aplicaciones personalizables.

1.6 Librería OpenCV.

Menciona (Arevalo, González, & Ambrosio, 2016) el 13 de junio del 2000, Intel® Corporation anunció que estaba trabajando con un grupo de reconocidos investigadores en visión por computador para realizar una nueva librería de estructuras/funciones en lenguaje C. Esta librería proporcionaría un marco de trabajo de nivel medio-alto que ayudaría al personal docente e investigador a desarrollar nuevas formas de interactuar con los ordenadores. Este anuncio tuvo lugar en la apertura del IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Había nacido The Open Computer Vision Library y lo hacía bajo licencia BSD (Software Libre).

La librería OpenCV es una API de aproximadamente 300 funciones escritas en lenguaje C que se caracterizan por lo siguiente:

- Su uso es libre tanto para su uso comercial como no comercial.
- No utiliza librerías numéricas externas, aunque puede hacer uso de alguna de ellas, si están disponibles, en tiempo de ejecución.
- Es compatible con The Intel® Processing Library (IPL) y utiliza The Intel® Integrated
- Performance Primitives (IPP) para mejorar su rendimiento, si están disponibles en el sistema.

1.6.1.- Estructura.

La librería OpenCV está dirigida fundamentalmente a la visión por computador en tiempo real. Entre sus muchas áreas de aplicación destacarían: interacción hombre máquina segmentación y reconocimiento de objetos; reconocimiento de gestos; seguimiento del movimiento; estructura del movimiento y robots móviles. La figura 1.8 muestra la estructura de la librería OpenCV empleada para el desarrollo del presente proyecto.

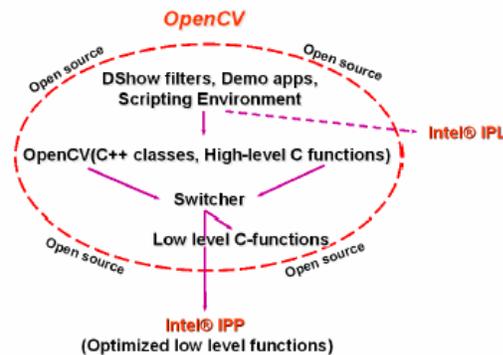


Figura 1.8. Estructura de la librería OpenCV.

Fuente. - (Arevalo, González, & Ambrosio, 2016)

1.7 Conversor de texto a voz.

Los llamados programas Text-To-Speech (TTS) son alternativas para lectura en voz alta de textos, tanto en equipos de cómputo portátiles o de escritorio. Ayuda a los usuarios a no requerir de la intervención humana para la lectura de un texto.

Los TTS están formados por dos módulos: el módulo de procesamiento de lenguaje natural (encargado de generar la transcripción fonética y las marcas prosódicas) y el módulo de procesamiento digital de la señal (encargado de generar la onda acústica).

La técnica más utilizada de generación de la onda acústica es la síntesis por concatenación, a pesar de que existen otras técnicas, como por ejemplo la síntesis por formantes y la síntesis articulatoria.

La técnica de síntesis por concatenación consiste en unir segmentos cortos pregrabados de un mismo locutor para formar locuciones. Todos los esfuerzos de los sistemas de concatenación están centrados en minimizar las discontinuidades espectrales y prosódicas en los puntos de concatenación. (Barrobés & Ruiz, 2017)

1.8 Equipo Reconocimiento óptico de caracteres (OCR).

El reconocimiento óptico de caracteres (Optical Character Recognition), es un proceso dirigido a la digitalización de textos, los cuales son identificados automáticamente a partir de una imagen que contiene símbolos o caracteres que pertenecen a un alfabeto determinado, para después de ser reconocidos como secuencias de dicho alfabeto y ser posteriormente almacenados en forma de datos.

El algoritmo K-NN (K vecinos más próximos), es un método utilizado para el reconocimiento de caracteres debido a su sencillez y efectividad. (Ortuño López, 2016)

1.9 Cámara web.

Para la captura de las imágenes que serán procesadas se utilizó una cámara web logitech C922 PRO HD STREAM (figura 1.9). Posee un sensor de cristal full HD con una resolución máxima de 1080p/30 fps – 720/60 fps. (logitech, 2018)

Tiene la ventaja de poseer un terminal USB que facilitará la integración con cualquier otro dispositivo en este caso particular con la tarjeta Raspberry Pi.



Figura 1.9. WebCam C922 PRO HD STREAM.

Fuente. - (logitech, 2018)

1.10.- Librería Espeak texto a voz.

Según (SOURCEFORGE, 2018) es un sintetizador de voz de software de código abierto compacto para inglés y otros idiomas, utiliza un método de "síntesis de formantes", permitiendo que muchos idiomas se proporcionen en un tamaño pequeño.

Posee las siguientes características:

- Incluye diferentes voces.
- Produce salida de voz como un archivo WAV
- Puede traducir el texto a códigos de fonemas, por lo que podría adaptarse como una interfaz para otro motor de síntesis de voz.
- No necesita de conexión a internet.

1.11.- Librería Python – Tesseract.

Python-Tesseract es una herramienta orientada al reconocimiento óptico de caracteres (OCR) para Python. Es decir, reconoce y lee el texto incrustado en las imágenes, puede leer todos los tipos de imágenes compatibles con Python Imaging Library, incluidos jpeg, png, gif, bmp, tiff y otros, mientras que Tesseract-Ocr de forma predeterminada solo admite tiff y bmp. Además, si se usa como script, Python-Tesseract imprimirá el texto reconocido en lugar de escribirlo en un archivo. (Foundation, 2018)

CAPÍTULO 2

MARCO METODOLÓGICO

En este capítulo se plantean los aspectos metodológicos y prácticos que se utilizaron para desarrollar un dispositivo que mediante visión artificial permita adquirir imágenes con palabras para la conversión a audio. Además, se detalla el tipo de investigación que se realizó y la metodología empleada.

2.1 Tipo y diseño de investigación

De acuerdo con los criterios expuestos por Palella y Martins (2006, p.103) definen dicha investigación como aquella que “conlleva a diseño o creaciones dirigidas a cubrir una necesidad y basadas en conocimientos anteriores. De allí que el término proyectivo refiere a un proyecto en cuanto aproximación o modelo teórico. El investigador puede llegar a está mediante vías diferentes, las cuales involucran procesos, enfoques, métodos y técnicas propias”

Expuestos por Hurtado, J (2010, p.117) la presente investigación fue Proyectiva por cuanto intenta “mediante este tipo de investigación se intenta proponer soluciones a una situación determinada a partir de un proceso previo de indagación. Implica explotar, describir y proponer alternativas de cambio, más no necesariamente ejecutar la propuesta”

Por lo anterior expuesto se puede decir que la investigación es de tipo proyectiva puesto que, a través de estudios previos, propone el desarrollo de un sistema conversor de imágenes con palabras a audio, brindando un aporte de gran importancia a la vanguardia tecnológica, el cual involucra procesos como enfoque, métodos y técnicas propias.

Bavaresco (2001 p.28), plantea que “La investigación también puede ser de campo; cuando se realiza en el propio sitio donde se encuentra el objeto de estudio. Ello permite el conocimiento más a fondo del problema por parte del investigador y puede manejar los datos con más seguridad. Así podrá soportarse en diseños exploratorios, descriptivos, experimentales y predictivos”.

Considerando las definiciones anteriores y basándose en las características del presente estudio, resulta aplicable el diseño No experimental, debido a que no se pretende modificar o manipular la realidad actual de las variables, es decir, se observan las situaciones ya existentes, sin ser provocadas intencionalmente por el investigador. Es transversal, puesto a que se recopilan los datos una sola vez en un momento determinado.

Por ultimo según la forma de obtener los datos es de Campo, debido a que se realiza en el propio sitio donde se encuentra. Cuando se conoce el tipo y diseño de investigación, se tiene de manera precisa y clara a donde pretende llegar esta y con qué finalidad se realiza este importante trabajo.

2.2 Técnicas e instrumentos de recolección de datos

Hace referencia al uso de diferentes técnicas e instrumentos que son utilizados para buscar información con la finalidad de manifestar el problema planteado. Toda investigación se sustenta en los datos y la información recolectada durante la elaboración de la misma. Esta documentación permite al investigador el dominio de cada uno de los procesos relacionados con el proyecto, lo cual es vital para la elaboración objetiva y respaldo de propuestas, conclusiones, recomendaciones, entre otras.

Ahora bien, para la realización de la investigación se utilizó como técnica e instrumento para recolectar datos, la observación de manuales técnicos, herramientas electrónicas, software, entre otras. En esta investigación se utilizó la técnica de la observación documental, de presentación resumida, descrita por Balestrini (2006, p. 152), como “una lectura general de los textos que contienen las fuentes de información que son de mucho interés, extrayendo los datos idénticos de utilidad para la investigación”.

La misma fue ejecutada en esta investigación a partir de las búsquedas especializadas de documentos, patentes, informes o publicaciones relativas a tendencias tecnológicas asociadas

con la visión artificial. En cuanto al instrumento aplicado, Arias (2006, p. 69) lo define como “cualquier recurso, dispositivo o formato (en papel o digital) que se utiliza para obtener, registrar o almacenar información.

2.3 Metodología seleccionada.

El desarrollo del sistema de conversión de texto a audio planteado tiene como objetivo la conversión de imágenes con texto a audio para la ayuda de personas invidentes. Según Jiménez, P. (2009) la metodología de diseño debe garantizar la reducción de errores, en el caso que se generen, sean detectados de manera anticipada; por ello, la fase de diseño está acompañada de puntos de verificación de las distintas etapas antes de iniciar el montaje del circuito electrónico.

Para la verificación del correcto funcionamiento del sistema es el simulador basado en un software de computadora. No se asumirá que el programa está diseñado correctamente sino hasta realizar las pruebas correspondientes y los resultados sean los esperados.

Para el desarrollo del sistema conversor texto a audio, se utilizará el lenguaje de programación Python 2.7 propietario de Raspberry Pi, versión libre para estudiantes que permite utilizar todas las herramientas necesarias. Para el diseño del circuito impreso se utilizará PROTEUS 8.

Para el éxito del desarrollo del proyecto, es importante que los resultados de la simulación sean correctos y no arrojen errores. Luego que los resultados de la simulación son los deseados, se procede a realizar el montaje de los diferentes elementos, se procede entonces a la verificación del funcionamiento mediante la toma de mediciones con el instrumento adecuado. Finalmente se procede al diseño de la tarjeta definitiva del circuito impreso.

Además, se utilizó la adaptación realizada por Jiménez, P. (2009), de la metodología planteada por Angulo, debido a la concordancia de los objetos con las fases propuestas por el autor. De este modo, la metodología utilizada es directa al acudir a las fases que el mismo propone.

En la primera fase se definirán las especificaciones; en la segunda se realizará la adaptación entre el hardware y el software; en la tercera, se realizará la codificación del programa; en la cuarta, se establecerá la implementación del hardware; en la quinta, se ejecutará la depuración

del software; en sexto lugar se llevará a cabo la implementación del hardware con el software. Y en séptimo lugar se construirá el modelo definitivo y se realizará las pruebas finales. Quedando constituidas según se expone a continuación:

Fase I. Definición de las especificaciones.

El presente sistema a implementarse tiene como fin la captación de imágenes con palabras mediante una cámara web, para la posterior conversión a audio mediante la utilización de una tarjeta Raspberry Pi. Y para la salida de audio se utilizó auriculares conectados directamente a la tarjeta por medio de un conector jack 3,5 mm.

Fase II. Adaptación entre hardware y software.

La comunicación entre la tarjeta y el resto de dispositivos se realizó por medio de cables pertenecientes a los mismos dispositivos como es el caso de la cámara se comunicará con la tarjeta mediante cable USB, el auricular por medio de un cable con conector jack 3,5 mm. Finalmente se utilizó una batería externa con un voltaje de 5 voltios y una corriente de 2,1 Amperios, conectados directamente a la tarjeta por medio del cable micro USB, como fuente de alimentación.

Fase III. Codificación del programa.

Para la codificación del programa se utilizó el lenguaje de programación compatible de Raspberry Pi, llamado Phyton, basado en la utilización de códigos de línea y librerías compatibles, que ayudaran a la conversión del texto a audio para los fines pertinentes del proyecto.

Fase IV. Implementación del hardware.

Se elaboró un circuito de control para la captura de las imágenes, tomando en cuenta las características de todos los materiales a utilizar y dependiendo de las necesidades que se presentaron en la fase de pruebas por parte del usuario se realizó el diseño definitivo de la parte de control.

Fase V. Depuración del software.

Luego de realizar las primeras pruebas de funcionamiento, se procedió a mejorar el algoritmo con el fin de tener un óptimo funcionamiento, en el ámbito de la captura de las imágenes en diferentes situaciones de iluminación y posición de la imagen.

Fase VI. Integración del hardware con el software.

Para un adecuado funcionamiento se realizó pruebas de las conexiones entre el hardware y el software para comprobar una correcta comunicación entre las dos partes, se probó el sistema en diferentes condiciones.

Fase VII. Construcción del modelo definitivo y pruebas finales.

En esta última fase, se realizó la implementación de un dispositivo final con el propósito de hacerlo portable y que tenga un correcto funcionamiento y de fácil manipulación.

CAPÍTULO 3

PROPUESTA

El presente capítulo expone la propuesta en base a lineamientos técnicos teóricos, para el desarrollo y uso de un dispositivo que mediante visión artificial permita adquirir imágenes con palabras para la conversión a audio, orientado a la ayuda de personas invidentes, el cual se fundamenta en la necesidad del acceso a la información de textos cortos de libros, revistas, etc.

En tal sentido, la propuesta se encuentra apoyada en los resultados obtenidos por la aplicación de los instrumentos de la investigación, los cuales reflejan la importancia que tiene para las personas invidentes, contar con un dispositivo portátil que permita la conversión de texto a audio, referidas para ayuda de las personas invidentes en la lectura de textos cortos, siguiendo especificaciones estrictas de ayuda a este tipo de personas, representados por la misión, visión y el conocimiento necesario para lograr un desarrollo óptimo de las tareas y actividades llevadas para el desarrollo de dicho proyecto.

3.1 Misión.

Introducción

La misión de la propuesta se dirige a establecer la conversión a audio de imágenes que contenga texto, responden al desarrollo de un sistema portátil, cuyo objetivo es garantizar la ayuda para la lectura de textos de personas invidentes. Todo ello con la finalidad de otorgar una alternativa de lectura de textos cortos orientado a las personas invidentes.

3.2 Justificación

El propósito de la propuesta al desarrollar un dispositivo que mediante visión artificial permita adquirir imágenes con palabras para la conversión a audio, orientado a la ayuda de personas invidentes, es generar a este tipo de personas, el acceso a la información que en muchos de los casos es escasa. Es decir, con la aplicación de la visión artificial, se crea el dispositivo que permite la conversión de texto corto a audio.

Dentro del ámbito social, la propuesta se justifica en la aplicación que tendrá dentro del sector de las personas invidentes al ser de simple utilización por estar formado por una cámara para la captura de las imágenes, por ser un dispositivo portátil que permitirá que el mismo sea trasladado de un lugar a otro.

En lo que se refiere al punto de vista económico se seleccionó dispositivos y materiales que en su mayoría son comerciales de costos accesibles.

3.3 Factibilidad de la propuesta.

El estudio de la factibilidad del presente proyecto “desarrollo de un dispositivo que mediante visión artificial permita adquirir imágenes con palabras para la conversión a audio, orientado a la ayuda de personas invidentes” ayuda a determinar la disponibilidad de los recursos visto desde los aspectos técnico, operativo y económico.

3.3.1 Estudio técnico.

El estudio técnico permite determinar si la propuesta tendrá éxitos en los procesos de diseño, implementación, funcionamiento y tecnológico. Para el éxito de este proyecto se debe considerar algunos aspectos teóricos y técnicos, que servirán de ayuda para la adecuada captura y posterior procesamiento de la imagen.

La selección de los componentes se basó en la versatilidad, facilidad y disponibilidad de los diferentes elementos en el mercado. A continuación, se detallan algunas características que fueron tomadas en cuenta para la elección de cada uno de ellos.

Características de la cámara.

Una de las tareas principales para el presente proyecto es la captación de imágenes, donde se requiere de una imagen de alta calidad, contraste y de un contorno definido.

A demás de estos parámetros se requiere contar con una cámara que sea económica, que trabaje con un valor de 640*480 pixeles como mínimo. En el mercado existen cámaras profesionales de marcas conocidas por su calidad de imágenes captadas, las mismas que son de muy alto costo que rodean los valores de 600 USD hasta los 1200 USD, lo cual impide la adquisición de las mismas.

Tomando en cuenta los parámetros anteriormente señalados, se procedió a la búsqueda de una cámara que se ajuste a las necesidades tanto técnicas como económicas. Con lo cual se pudo recolectar información con la finalidad de generar como resultado un cuadro comparativo (tabla 3.1) de las diferentes cámaras que se ajustan a las necesidades de este proyecto.

Tabla 3.1. Cuadro comparativo de las cámaras web consideradas.

IDENTIFICACIÓN	RESOLUCIÓN (Pixeles)	SENSOR	VELOCIDAD DE TRANSMISIÓN (cuadro/seg)	PIXELES	PRECIO (USD)
Lifecam Studio HD 1080p	19 MP	CMOS	60	2560*2048	139,99
Logitech C922 Full HD	15 MP	CMOS	60	1920*1080	130,00
Logitech C270	3 MP	CMOS	30	1280*720	31,99
Genius Facecam 320x	VGA	CMOS	15	640*480	18,00
Genius Facecam 1000x	1,3 MP	CMOS	30	1280*720	25,00

Fuente. – Elaborado por el autor.

Luego de realizar una comparación entre las diferentes webcams presentes en el mercado, se optó por realizar la adquisición de la cámara web Logitech C922 PRO, al cumplir con los requerimientos necesarios para realizar una adecuada captura de la imagen, además de contar

con enfoque automático HD, de esta manera se podrá adquirir imágenes nítidas de alta definición. Dichas especificaciones técnicas dadas por el fabricante se muestran en la tabla 3.2.

Tabla 3.2. Características técnicas de la cámara web.

CARACTERÍSTICAS TÉCNICAS	
Ata resolución Cristal Full HD, sensor de color.	
Resolución: 15 M pixeles (Enfoque automático)	
Formato de video: 24 bit Color verdadero	
Velocidad de transmisión:	
1080p	30 fps
720p	60 fps
Capo visual de 78 grados	
Balance de blancos: Automático.	

Fuente. – Elaborado por el autor.

Software para el reconocimiento óptico de caracteres.

A continuación, se realiza una comparación de las herramientas disponibles para el proceso de reconocimiento óptico de caracteres, tomando en cuenta las características de cada una de ellas se procederá a elegir la mejor opción para la implementación del algoritmo, como se muestra en la tabla 3.3.

Tabla 3.3. Software disponible para el reconocimiento óptico de caracteres.

SOFTWARE	LICENCIA	TIEMPO DE RESPUESTA	CALIDAD DE LAS IMÁGENES A SER TRATADAS	CONECTIVIDAD A INTERNET	SOPORTA EL IDIOMA ESPAÑOL
Tesseract	Gratuita	Bueno	Buena	No	Si
Gocr	Gratuita	Bueno	Excelente	No	No
Ocrad	Gratuita	Bueno	Excelente	No	Si

Tocr	Comercial	Excelente	Buena	Si	Si
Leadtools OCR SDK	Comercial	Excelente	Excelente	Si	Si

Fuente. – Elaborado por el autor.

Según las características señaladas en la tabla 3.3, el software Tesseract es el más viable para el desarrollo del algoritmo ya que reúne las siguientes características:

- Es un software libre.
- Tiene un porcentaje de acierto por encima del 95%.
- No necesita una imagen con alta densidad de pixeles.
- Soporta varios idiomas, principalmente el español
- No requiere de conexión a internet.

Tesseract tiene un data o tessdata que contiene 20 muestras de 94 caracteres en 5 diferentes fuentes, puede ser ampliado mediante entrenamiento, lo que permite crear nuevas definiciones de letras. Tiene soporte para utf-8 y puede reconocer más de 100 idiomas y también admite varios formatos de salida: texto plano, pdf, tvs y html.

Además, los softwares Tocr y Leadtools OCR SDK, son softwares comerciales que dentro del mercado tiene un costo aproximado de 200 dólares americanos, ya que cuentan con características de tiempo de respuesta menores y poco error en el reconocimiento de caracteres.

Software para la extracción de caracteres.

El algoritmo implementado usa la función de OpenCV llamada threshold, (Gutierrez, 2019), lo define como una función que consiste en separar regiones de una imagen correspondientes a los objetos que se desea analizar. Dicha separación se basa en la variación de la intensidad entre los pixeles de los objetos y los pixeles de fondo.

Para diferenciar los pixeles que se deben extraer del resto, se realizara una comparación de cada valor de intensidad con respecto a un umbral, se establecerá un valor determinado para identificar (asignar un valor de 0 para negro y 255 para blanco), o un valor que se adapte a sus necesidades.

Umbral binario.

Para ilustrar como funciona se considerara una imagen de origen con valor de intensidad de pixeles $src(x,y)$, como se muestra en la figura 3.1, donde la línea azul representa el umbral fijo $thresh$.

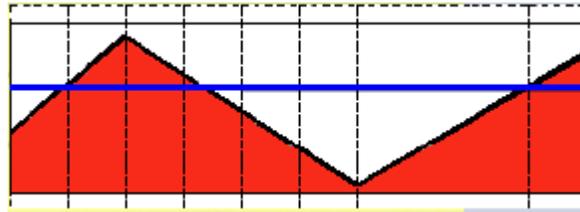


Figura 3.1. Umbral fijo, función thresh.

Fuente. - (Gutierrez, 2019)

La operación de umbral binario se expresa en la ecuación 1:

$$dts(x,y) = \begin{cases} maxVal & \text{if } src(x,y) > thresh \\ 0 & \text{otherwise} \end{cases}$$

Ecuación 1. Modelamiento matemático de la función thresh, binario.

Fuente. - (Gutierrez, 2019)

Por lo tanto, si la intensidad del pixel $src(x,y)$ es superior al $thresh$, entonces la nueva intensidad de los pixeles se establece en un $MaxVal$. De lo contrario, los pixeles se establecen en 0. Como se muestra en la figura 3.2.

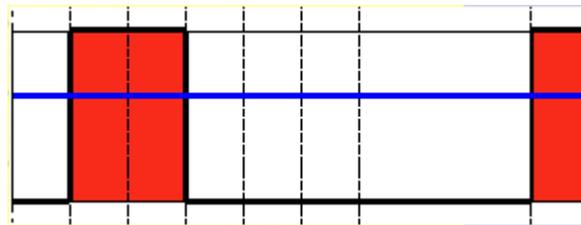


Figura 3.2. Umbral binario, función thresh.

Fuente. - (Gutierrez, 2019)

En la actualidad existen algunos métodos para la modificación de imágenes uno de estos métodos es el uso de los filtros, ya sea para la detección de bordes, para la modificación del aspecto y para la eliminación de ruido.

Uno de estos filtros el Gaussiano, que usa el principio de promediar los pesos de los valores vecinos a ambos lados de un pixel, este tipo de filtro tiene el inconveniente del difuminado de

los bordes de una imagen, al reducir el ruido gaussiano produce pequeñas variaciones en la imagen y es sensible a otros ruidos presentes en la digitalización de las imágenes.

Tomando en cuenta dichos inconvenientes que presenta el filtro de gauss, se optó por usar el método del umbral binario ya que es un método simple y no consume recursos en el procesamiento de la imagen que afectan directamente al rendimiento de la tarjeta a usarse para el presente proyecto.

Función Tesseract

Para la extracción de caracteres se utilizará la función Tesseract, esta función tiene algunas configuraciones que dependiendo del caso de aplicación serán de ayuda para el propósito del presente proyecto.

- La primera configuración que se debe realizar es el lenguaje (*lang*), si no se especifica ninguno, la función asumirá el idioma inglés.
- Se puede configurar a Tesseract para que solo ejecute un subconjunto de análisis de diseño y asuma cierta forma de imagen. Las opciones para (*psm*) son:
 - 0: Sólo orientación y detección de script (OSD).
 - 1: Segmentación automática de páginas con OSD.
 - 2: Segmentación automática de páginas, pero sin OSD u OCR.
 - 3: Segmentación de página completamente automática, pero sin OSD.
(Defecto)
 - 4: Supone una sola fila de texto de tamaños variables.
 - 5: Supone un solo bloque uniforme de texto alineado verticalmente.
 - 6: Supone un solo bloque de texto uniforme.
 - 7: Trata a la imagen como una sola línea de texto.
 - 8: Trata a la imagen como una sola palabra.
 - 9: Trata a la imagen como una sola palabra en un círculo.
 - 10: Trata a la imagen como un solo carácter.
- Además, se puede especificar el modo de motor OCR (*oem*) cuyas opciones son:
 - 0: Tesseract original solamente.
 - 1: Sólo redes neuronales LSTM.

- 2: Tesseract + LSTM.
- 3: Predeterminado, basado en lo que está disponible.

Tomando en cuenta las configuraciones anteriores se determinó que el idioma será español, el modo psm será 4 y el motor oem será 3.

Software espeak para convertir texto a voz.

Es un programa que sirve para transformar texto en voz, su principal característica es que la voz producida no es muy realista, suena muy parecida a un robot. El software permite modificar algunas características de la voz como lo es, el idioma, la velocidad, el tono y la amplitud de la voz, como se muestra en la tabla 3.4. Para el propósito de este proyecto se procedió a realizar los siguientes ajustes al software espeak:

- Se procedió a indicar el idioma, ya que el idioma por defecto del software es el inglés.
- Se debe configurar el tipo de voz a usarse, espeak cuenta con varias alternativas disponibles, entre ellas la voz de hombre (m) y mujer (f).
- Finalmente se configuro la velocidad del habla, se utilizó la opción de 160 palabras por minuto.

Tabla 3.4. Descripción del comando espeak

Opciones de comando	Descripción
Espeak	El comando para ejecutar el programa empieza con eSpeak
-v<voz>	Sirve para seleccionar el lenguaje por ejemplo -v es-la (español latino), -v es (español).
-s	Indica la velocidad de lectura, El valor, por defecto, es 160.
texto	Es el texto que va ser leído, debe ir entre comillas.

Fuente. – Elaborado por el autor.

Una vez efectuadas estas configuraciones se procedió a realizar pruebas en el ámbito de entendimiento y la velocidad de pronunciación de la voz, enfocada a la lectura de un texto corto. Se configuro con tres tipos de velocidad de lectura, el primero por defecto; es decir 160 palabras por minuto, el segundo con un valor menor de 120 y el tercero con un valor superior de 180. También se configuro el algoritmo para que la voz emitida sea tanto de hombre como de mujer.

Para la ejecución de dichas pruebas se procedió a realizar una encuesta de 10 preguntas a 3 personas, en el ámbito de comprensión de lectura del texto realizada por el algoritmo y también en la apreciación de las personas sobre las diferentes configuraciones como lo son: el tipo de voz, velocidad de pronunciación, y pausa entre palabras.

Como primera parte de este apartado se presentó el siguiente texto, para ser leído por parte del software espeak:

“La fascinante historia del Rey León. En la selva virgen, donde los animales salvajes viven y luchan manteniendo el equilibrio natural que impone la ley del más fuerte, el león Mufasa reina junto a su esposa Saraby. Ambos han traído al mundo a Simba. Simba es sucesor al trono, algo que no le gusta a su tío Scar, con la ayuda de tres malvadas y tontas hienas, Scar urde una treta en la que su hermano y rey Mufasa muere en una estampida y provoca que Simba crea que ha sido por su culpa y decide huir a la selva, después de que las tres hienas quisieran matarlo también. Allí conoce a un suricato llamado Timón y a un facóquero llamado Pumba, que le adoptaran y, además de entablar amistad, le enseñan la filosofía de vivir sin preocupaciones: el Hakuna Matata.”

Como segundo punto, se procedió a realizar las siguientes preguntas con una configuración de velocidad de reproducción de 160 palabras por minuto (valor por defecto), con las respectivas respuestas por parte de los encuestados, como se muestra en la tabla 3.5.

Tabla 3.5. Primera encuesta.

1. Desde su punto de vista. ¿La voz emitida por el dispositivo es clara y comprensible?			
SI	3 Personas	NO	Ninguna
2. Desde su punto de vista. ¿La pronunciación del texto, fue la correcta?			
SI	2 Personas	NO	1 Persona

3. ¿Considera usted que la pausa entre palabras es la adecuada?							
SI	3 Personas			NO	Ninguna		
4. ¿Cuál fue su percepción de la velocidad de lectura del texto propuesto?							
Muy rápido	Ninguna	Rápido	Ninguna	Normal	3 Personas	Lento	Ninguna
5. ¿Preferiría escuchar una voz masculina, o una voz femenina?							
Masculina	1 Persona			Femenina	2 Personas		
6. ¿La voz sintetizada le pareció robotizada?							
SI	2 Personas			NO	1 Persona		
7. ¿Considera usted que el texto hace referencia a la historia del rey león?							
Verdadero	3 Personas			Falso	Ninguna		
8. ¿El personaje principal de la historia propuesta es un león?							
Verdadero	3 Personas			Falso	Ninguna		
9. ¿El tío del personaje principal se llama Scar?							
Verdadero	2 Personas			Falso	1 Persona		
10. ¿El personaje de la historia propuesta conoce a dos animales llamados timón y pumba?							
Verdadero	2 Personas			Falso	1 Persona		

Fuente. – Elaborado por el autor.

Luego de un análisis de las respuestas proporcionadas por las personas encuestadas, se determinó que este tipo de configuración del software es el ideal para los fines del presente proyecto. Debido a que en la primera parte de la encuesta que hace referencia al entendimiento de lectura de la voz sintetizada por parte del algoritmo diseñado, las respuestas reflejan que la mayoría de las personas encuestadas consideran que la voz emitida es clara y comprensible, además que la velocidad de pronunciación es la correcta y con una pausa adecuada entre palabras. Finalmente, como segunda parte de la encuesta esta se enfoca en la comprensión de del texto, mediante preguntas enfocadas al contenido del texto leído, las cuales fueron respondidas correctamente, considerando que dichas configuraciones son las adecuadas.

A continuación, se procedió a realizar la misma encuesta, con un valor de velocidad mayor al valor por defecto de la voz sintetizada, para este caso se consideró un valor de 180, obteniendo los siguientes resultados por parte de los encuestados, como se muestra en la tabla 3.6.

Tabla 3.6. Segunda encuesta.

1. Desde su punto de vista. ¿La voz emitida por el dispositivo es clara y comprensible?							
SI	1 Persona			NO	2 Personas		
2. Desde su punto de vista. ¿La pronunciación del texto, fue la correcta?							
SI	Ninguna			NO	3 Personas		
3. ¿Considera usted que la pausa entre palabras es la adecuada?							
SI	Ninguna			NO	3 Personas		
4. ¿Cuál fue su percepción de la velocidad de lectura del texto propuesto?							
Muy rápido	3 Personas	Rápido	Ninguna	Normal	Ninguna	Lento	Ninguna
5. ¿Preferiría escuchar una voz masculina, o una voz femenina?							
Masculina	Ninguna			Femenina	3 Personas		
6. ¿La voz sintetizada le pareció robotizada?							
SI	3 Personas			NO	Ninguna		
7. ¿Considera usted que el texto hace referencia a la historia del rey león?							
Verdadero	1 Persona			Falso	2 Personas		
8. ¿El personaje principal de la historia propuesta es un león?							
Verdadero	1 Persona			Falso	2 Personas		
9. ¿El tío del personaje principal se llama Scar?							
Verdadero	Ninguna			Falso	3 Personas		
10. ¿El personaje de la historia propuesta conoce a dos animales llamados tímón y pumba?							
Verdadero	Ninguna			Falso	3 Personas		

Fuente. – Elaborado por el autor.

Una vez analizadas las respuestas proporcionadas por las personas estas consideran que la voz emitida no es clara impidiendo de esta manera el entendimiento del texto leído, de igual manera consideran que la pronunciación no es la adecuada y la pausa entre palabras es muy corta. En el ámbito de comprensión de lectura del texto propuesto, la mayoría de personas encuestadas no supo responder correctamente, es decir no logro comprender la lectura. Se concluye entonces, que dicha velocidad propuesta no será aplicada para el presente proyecto, ya que la mayoría de los encuestados no logro comprender el texto leído por el software.

Finalmente se procedió a realizar la misma encuesta, con un valor de velocidad de reproducción menor de la voz emitida, dicho valor será de 120 palabras por minuto. Obteniendo las siguientes respuestas por parte de las personas encuestadas, tabla 3.7.

Tabla 3.7. Tercera encuesta.

1. Desde su punto de vista. ¿La voz emitida por el dispositivo es clara y comprensible?							
SI	2 Personas			NO	1 Persona		
2. Desde su punto de vista. ¿La pronunciación del texto, fue la correcta?							
SI	Ninguna			NO	3 Personas		
3. ¿Considera usted que la pausa entre palabras es la adecuada?							
SI	Ninguna			NO	3 Personas		
4. ¿Cuál fue su percepción de la velocidad de lectura del texto propuesto?							
Rápido	Ninguna	Normal	Ninguna	Lento	Ninguna	Muy Lento	3 Personas
5. ¿Preferiría escuchar una voz masculina, o una voz femenina?							
Masculina	1 Persona			Femenina	2 Personas		
6. ¿La voz sintetizada le pareció robotizada?							
SI	3 Personas			NO	Ninguna		
7. ¿Considera usted que el texto hace referencia a la historia del rey león?							
Verdadero	3 Personas			Falso	Ninguna		
8. ¿El personaje principal de la historia propuesta es un león?							
Verdadero	3 Personas			Falso	Ninguna		

9. ¿El tío del personaje principal se llama Scar?			
Verdadero	3 Personas	Falso	Ninguna
10. ¿El personaje de la historia propuesta conoce a dos animales llamados timón y pumba?			
Verdadero	3 Personas	Falso	Ninguna

Fuente. – Elaborado por el autor.

Una vez analizadas las respuestas proporcionadas por las personas encuestadas, se determinó que, a pesar de que las respuestas fueron acertadas en la parte de comprensión de lectura, no es recomendable el uso de esta configuración de velocidad, ya que la mayoría de personas consideran que dicha velocidad es muy lenta y afecta directamente al tiempo de reproducción de la voz emitida.

3.3.2 Características de los componentes.

Fuente de alimentación: Según las especificaciones dadas por el fabricante, para un adecuado funcionamiento de la tarjeta Raspberry Pi, necesita un voltaje de 5 voltios y una corriente mínima de 2.5 A.

Se utilizó un power bank que servirá de alimentación para la tarjeta Raspberry Pi, así como a los diferentes dispositivos electrónicos que conforman el sistema los mismos que son: la cámara web y el pulsador.

El power bank posee las siguientes características:

- Voltaje de salida: 5 Voltios
- Intensidad de salida: 2,1 Amperios
- Capacidad: 2000 mAh
- Dimensiones: 114×63×24 mm
- Peso: 445 gramos.

Para la adquisición de la batería se consideró un parámetro muy importante que es la potencia, este factor definirá el adecuado funcionamiento de todo el sistema.

La batería a utilizar tiene una capacidad de 2200 mAh, que según la ecuación de la capacidad se obtendrá el tiempo promedio que la fuente externa permitirá el uso del sistema completo.

$$C = x \times t$$

Ecuación 2. Ecuación para el cálculo de la capacidad de una batería.

Fuente. - (AutoSolar, 2015)

Donde:

- x: Amperios que requiere la tarjeta. (Amperios)
- t: Tiempo aproximado de duración de la fuente externa. (horas)
- C: Capacidad de la fuente externa. (Amperios / horas)

Empleando la ecuación 2, se procederá a calcular el tiempo de duración de la fuente externa. Empleando los valores de corriente y capacidad antes mencionados.

$$t = \frac{C}{x}$$

$$t = \frac{2 \text{ Ah}}{2,5 \text{ A}}$$

$$t = 0,8 \text{ horas.}$$

Según el cálculo anterior se puede determinar, que el tiempo aproximado de funcionamiento del sistema completo es aproximadamente 1 hora.

Dispositivo periférico: Hace referencia al pulsador, que es el responsable de capturar la imagen de interés.

La imagen a ser procesada se realizará mediante un pulsador, figura 3.3, el cual debe garantizar una adecuada captura, es por tal motivo que se eligió el diseño de un pulsador, a base de una resistencia en configuración pull down.

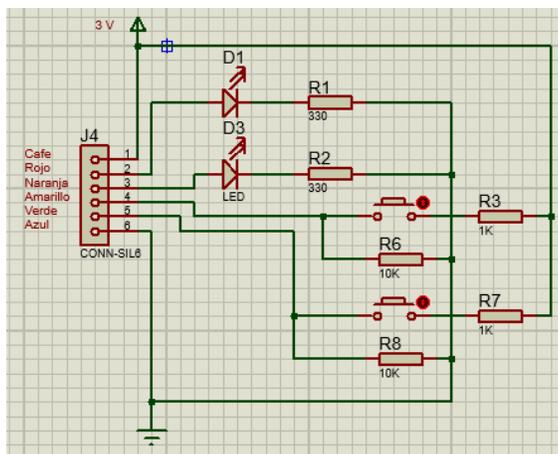


Figura. 3.3. Circuito para la captura de imágenes.

Fuente. - Elaborado por el autor.

El desarrollo del cálculo de los valores de las resistencias a utilizarse se basó, en los valores o datos obtenidos de la tarjeta Raspberry Pi.

La ecuación 3, hace referencia al cálculo del valor de la resistencia pull – down máxima.

$$R_{1 \text{ pull down max}} = \frac{V_{il}}{I_{l1}} = \frac{\text{Baja tensión de entrada}}{\text{Corriente de fuga de entrada}}$$

Ecuación 3. Ecuación para el cálculo del valor de la resistencia pull - down máxima.

Fuente. - (García , 2018)

A continuación, se procederá a realizar los cálculos respectivos.

$$R_{1 \text{ pull down max}} = \frac{0.6 \text{ V}}{5\mu\text{A}}$$

$$R_{1 \text{ pull down max}} = 120 \text{ K}\Omega$$

De igual manera la ecuación 4, hace referencia al cálculo del valor de la resistencia pull – down mínima.

$$R_{1 \text{ pull down min}} = \frac{V_{OH}}{I_{OH}} = \frac{\text{Salida de voltaje en alto}}{\text{Salida de corriente en alto}}$$

Ecuación 4. Ecuación para el cálculo del valor de la resistencia pull-up mínima.

Fuente. - (García , 2018)

A continuación, se procede a realizar el cálculo de dicha resistencia.

$$R_{1 \text{ pull down min}} = \frac{1.6 \text{ V}}{2 \text{ mA}}$$

$$R_{1 \text{ pull down min}} = 800 \Omega$$

De acuerdo a los valores obtenidos en los cálculos anteriores, se determina que el intervalo de las resistencias pull down es: $800 \Omega \leq 120 \text{ K}\Omega$. Por lo que los valores comerciales de las resistencias serian de $1\text{K}\Omega$, $10 \text{ K}\Omega$ respectivamente.

Cámara web: Encargada de capturar la imagen para su posterior procesamiento y reconocimiento de caracteres.

3.3.3 Parámetros de la cámara.

Campo de visión: Para (Sierra Álvarez, 2012), es el área del objeto que se captara por el sensor. Este se toma midiendo el área total de las plantillas, la distancia horizontal y vertical.

Para efectos de este proyecto se capturará documentos en formato A4 cuyas dimensiones son:

$$A4 = 21,0 \times 29,7 \text{ cm}$$

$$\text{Ancho} = 29,7 \text{ cm}$$

$$\text{Alto} = 21,0 \text{ cm}$$

Empleando la ecuación 5, que muestra el cálculo del área de un rectángulo se obtiene el siguiente valor:

$$\text{Área} = \text{ancho} \times \text{alto}$$

Ecuación 5. Ecuación para el cálculo del área de un rectángulo.

Fuente. - (Ecuador, 2011)

$$\text{Área} = 29,7 \times 21,0$$

$$\text{Área} = 623,7 \text{ cm}^2$$

Distancia de trabajo: La figura 3.4, muestra la distancia de trabajo, que se conoce como la distancia medida desde el lente de la cámara hasta el objeto.

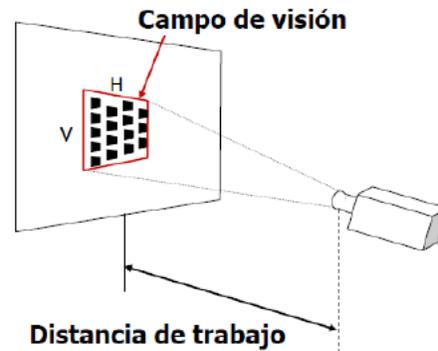


Figura 3.4. Campo de visión y distancia de trabajo.

Fuente. - (Sierra Álvarez, 2012)

Distancia focal: Sirve para conocer cuál será el ángulo visual, es decir, cuánto de la escena que se tiene en frente se puede capturar.

Existe una relación que se debe tomar en cuenta: Mientras más larga la longitud focal, menor será el ángulo visual que la lente puede capturar, y el aumento será mayor. Por el contrario, mientras menor sea la longitud, mayor será el ángulo y menor será el aumento.

En el caso de la cámara web, según las especificaciones obtenidas de los fabricantes, los valores tanto del ángulo visual como la longitud focal son:

- Ángulo visual: 78°
- Intervalo de la longitud focal: 5 – 25 mm

Según la ecuación 6, del campo visual o FOV (Field of view), viene expresada por un ángulo llamado ángulo de visión que se denota con la letra griega α .

$$\alpha = 2 \tan^{-1} \frac{d}{2f}$$

Ecuación 6. Ecuación del campo visual (FOV)

Fuente. - (Latin Fotografía, 2018)

Donde:

- α : Ángulo de visión.
- d: diagonal del sensor.
- f: distancia focal.

Teniendo en cuenta los valores del ángulo visual como la distancia focal, de la cámara a emplearse para este proyecto, el valor de la diagonal del sensor es:

Despejando d , de la ecuación 6, se obtiene la siguiente expresión:

$$d = 4f \tan \alpha$$

$$d = 4(5mm) \tan(78^\circ)$$

$$d = 94,09 \text{ mm}$$

Resolución de la imagen: Es el grado de detalle o mejor conocida como la calidad de una imagen digital, dicho valor generalmente se expresa en ppp (píxeles por pulgada). Cuantos más píxeles contenga una imagen por pulgada lineal, mayor calidad tendrá.

Las cámaras digitales tienen una calidad que se expresa en MegaPíxeles (MP). Considerando esta característica una cámara de 19 MP, es capaz de captar una fotografía con 19 millones de píxeles. Es así que el valor equivalente de un píxel en centímetros es: 1 píxel = 0,026458 centímetro [cm]. Según las especificaciones dadas por el fabricante la cámara que se utilizó para el desarrollo de este proyecto tiene la capacidad de captar imágenes con una resolución de 1920×1080 (ancho \times alto) píxeles.

Considerando dichos valores, y empleado el método de la regla de tres, se obtendrá una imagen con las siguientes características:

$$1 \text{ pixel} \rightarrow 0,026458 \text{ cm}$$

$$1920 \text{ pixeles} \rightarrow x$$

$$x = \frac{1920 \text{ pixeles} \times 0,026458 \text{ cm}}{1 \text{ pixel}}$$

$$x = 50,80 \text{ cm}$$

$$1080 \text{ pixeles} \rightarrow x$$

$$x = \frac{1080 \text{ pixeles} \times 0,026458 \text{ cm}}{1 \text{ pixel}}$$

$$x = 28,57 \text{ cm}$$

Obteniendo como resultado una imagen de las dimensiones de $50,80 \text{ cm} \times 28,57 \text{ cm}$.

Dentro del proceso de tratamiento de la imagen, se vio en la necesidad de resaltar una región de interés dentro del algoritmo. Esta región tiene una medida de 640x480 pixeles, es decir que el algoritmo diseñado se enfocara en realizar el análisis del texto que se encuentre dentro de dicha región. Según los valores anteriormente señalados, la nueva dimensión de la imagen es:

$$\begin{aligned}
 1 \text{ pixel} &\rightarrow 0,026458 \text{ cm} \\
 640 \text{ pixeles} &\rightarrow x \\
 x &= \frac{640 \text{ pixeles} \times 0,026458 \text{ cm}}{1 \text{ pixel}} \\
 x &= 16,93 \text{ cm} \\
 480 \text{ pixeles} &\rightarrow x \\
 x &= \frac{480 \text{ pixeles} \times 0,026458 \text{ cm}}{1 \text{ pixel}} \\
 x &= 12,70 \text{ cm}
 \end{aligned}$$

Obteniendo como resultado una imagen de las dimensiones de 16,93 cm × 12,70 cm.

Una vez realizado el análisis de estas variables se concluyó que:

- La selección del software para el desarrollo del proyecto se lo realizo basado en las facilidades de uso, y en los resultados obtenidos antes de la implementación.
- Se utilizará la tecnología disponible y se dará prioridad a los componentes que serán de fácil adquisición.
- Para el desarrollo del proyecto se contó con documentos que sirven como fuente de consulta.

3.4 Estudio económico.

Se hace referencia a los costos y beneficios que, al ser un proyecto de carácter social, estos serán intangibles. En las tablas 3.8, 3.9, 3.10 y 3.11, se detallan los recursos y componentes, así como el valor de cada uno que se necesitó para el desarrollo de “un dispositivo que mediante visión artificial permita adquirir imágenes con palabras para la conversión a audio, orientado a la ayuda de personas invidentes”.

Tabla 3.8. Recursos tecnológicos (SOFTWARE)

CANTIDAD	DESCRIPCIÓN	VALOR UNITARIO	VALOR TOTAL
1	PROTEUS 8.6 SP6 LABCENTER ELECTRONICS	0,00	0,00
1	NOOBS	0,00	0,00
1	Librería Tesseract	0,00	0,00
1	Librería Espeak	0,00	0,00
		TOTAL (USD)	0,00

Fuente. - Elaborado por el autor.

Tabla 3.9. Recursos tecnológicos (COMPONENTES ELECTRÓNICOS)

CANTIDAD	DESCRIPCIÓN	VALOR UNITARIO	VALOR TOTAL
2	Resistencia 1K Ω - ¼ W – 5%	0,02	0,04
2	Resistencia 10 K Ω - ¼ W – 5%	0,02	0,04
2	Resistencia 330 Ω - ¼ W – 5%	0,02	0,04
2	Diodo LED (Verde, Azul)	0,10	0,20
1	Placa perforada 7x9 (cm)	1,00	1,00
1	Tarjeta Raspberry PI modelo 3 B	80,00	80,00
1	Cámara web C922 PRO HD STREAM	130,00	130,00
1	Tarjeta micro SD Sandisk , 16 GB, clase 10	14,00	14,00
1	Power Bank 2000 mAh	15,00	15,00
1	Case para Raspberry	8,00	8,00
1	Convertidor HDMI a VGA	6,50	6,50
1	Auricular	5,00	5,00
2	Pulsador	0,25	0,50
		TOTAL (USD)	260,32

Fuente. - Elaborado por el autor.

Tabla 3.10. Componentes adicionales.

CANTIDAD	DESCRIPCIÓN	VALOR UNITARIO	VALOR TOTAL
1	Gorra	22,00	22,00
		TOTAL (USD)	22,00

Fuente. - Elaborado por el autor.

Tabla 3.11. Mano de obra.

CANTIDAD	DESCRIPCIÓN	H/H	VALOR UNITARIO	VALOR TOTAL
1	Código de programación	155	1,61	185,15
1	Implementación	30	1,61	48,30
			TOTAL (USD)	233,45

Fuente. - Elaborado por el autor.

Tabla 3.12. Valor total del proyecto.

CANTIDAD	DESCRIPCIÓN	VALOR UNITARIO	VALOR TOTAL
1	Recursos tecnológicos (software)	0,00	0,00
1	Componentes electrónicos	260,32	260,32
1	Componentes adicionales	22,00	22,00
1	Mano de obra	233,45	233,45
		TOTAL (USD)	515,77

Fuente. - Elaborado por el autor.

La tabla 3.12 muestra el costo total necesario para la implementación del proyecto fue de 515,77 dólares americanos. Los costos más altos para su desarrollo fueron: la mano de obra y los componentes electrónicos, siendo el 95,73% del total del proyecto.

3.5 Conclusión de la propuesta.

Luego de ser realizado el estudio antes mencionado, el proyecto “desarrollo de un dispositivo que mediante visión artificial permita adquirir imágenes con palabras para la conversión a audio, orientado a la ayuda de personas invidentes”, fue considerablemente bajo, por lo que se pudo desarrollar sin ningún inconveniente. Cabe mencionar que el autor fue el encargado de realizar las actividades de investigación, diseño, implementación, pruebas y puesta en marcha de proyecto y que dichos costos se tomaron en cuenta en el valor del código de programación.

A la vez se consideró que la propuesta fue viable, pues se contó con los recursos tecnológicos y económicos para la adquisición de los diferentes componentes del sistema, ya que los mismos actualmente se encuentran disponibles en el mercado

CAPÍTULO 4

IMPLEMENTACIÓN

El presente capítulo hace referencia a la implementación del sistema de conversión de texto a audio, donde se muestra las diferentes etapas que conforman el sistema. Además del circuito empleado para la captura de la imagen a ser tratada y los diagramas de flujo que hacen referencia a los sub procesos que realiza el algoritmo diseñado para el desarrollo del presente proyecto.

La tarjeta Raspberry Pi, es la encargada de realizar los diferentes procesos del tratamiento de imágenes, dichas etapas que conforman la adquisición y posterior tratamiento de las imágenes capturadas son las siguientes:

4.1 Desarrollo.

4.1.1 Proceso de instalación del sistema operativo Raspbian en la tarjeta microSD.

Como primer paso, se procedió a instalar el sistema operativo en la tarjeta microSD que se usara como memoria. Actualmente este proceso es sencillo.

Para (Llamas, 2018) NOOBS (New Out Of Box Software) es un instalador de sistemas operativos para Raspberry Pi. A continuación, se muestra los pasos a considerar para la adecuada instalación del sistema operativo en la tarjeta microSD.

- **Descargar NOOBS:** En primer lugar, se descargó Noobs desde la página web de Raspberry Pi. Cabe mencionar que se debe descargar la versión completa ya que la misma contiene el software Raspbian, que es el sistema operativo que se instaló, como se muestra en la figura 4.1.



Figura 4.1. Gestor de instalación de sistemas operativos Noobs.

Fuente. - (Llamas, 2018)

- Formatear la tarjeta microSD: En la figura 4.2, se muestra el proceso de formateo de la tarjeta micro SD, se recomienda utilizar una herramienta para formatear tarjetas microSD, para este caso se utilizó SD Card Formatter que está disponible de forma gratuita. Es aconsejable realizar el formateo de la tarjeta microSD para no tener errores de software en el futuro.

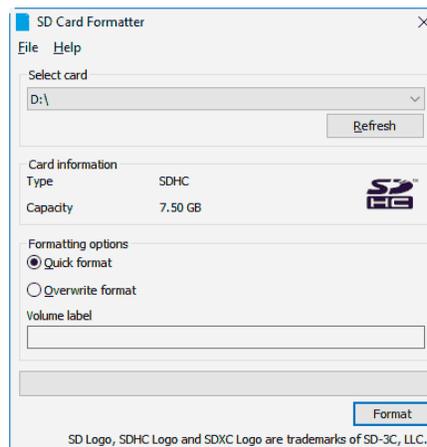


Figura 4.2. Cuadro de dialogo SD Card Formatter.

Fuente. - (Llamas, 2018)

- Instalar Noobs: Se descomprime el fichero antes descargado, una vez que el fichero termine de copiarse en su totalidad en la tarjeta microSD. Se procedió a insertar en la tarjeta Raspberry Pi.
- Instalar Raspbian: Una vez conectada la tarjeta Raspberry Pi, a la toma de corriente, al iniciar aparecerá un menú con los distintos sistemas operativos disponibles, se eligió el sistema operativo Raspbian, como se muestra en la figura 4.3.

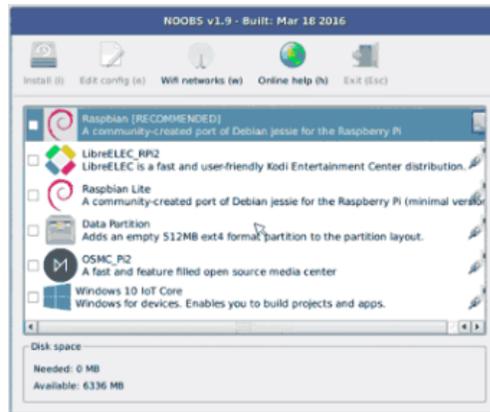


Figura 4.3. Elección del sistema operativo Raspbian.

Fuente. - (Llamas, 2018)

- La figura 4.4, muestra el proceso de instalación del sistema operativo Raspbian, en la tarjeta de memoria micro SD.



Figura 4.4. Proceso de instalación del sistema operativo Raspbian.

Fuente. - (Llamas, 2018)

- Una vez que el sistema operativo termina de instalarse por completo, es necesario reiniciar la tarjeta Raspberry Pi, para que el sistema operativo cargue, como se muestra en la figura 4.5.

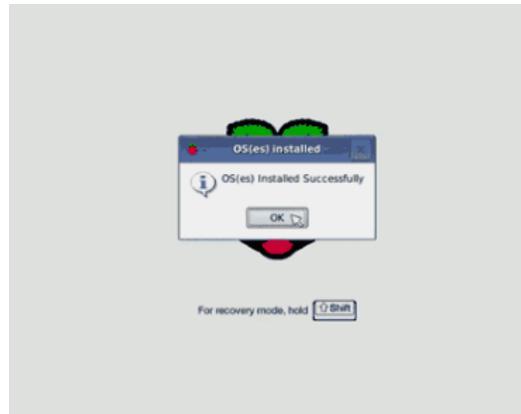


Figura 4.5. Finalización de la instalación del sistema operativo Raspbian.

Fuente. - (Llamas, 2018)

- Finalmente, la tarjeta Raspberry Pi se reinicia y ejecuta el sistema operativo instalado, siendo esta la parte final del proceso de instalación, figura 4.6.



Figura 4.6. Pantalla de inicio del sistema operativo Raspbian.

Fuente. - (Llamas, 2018)

4.1.2 Etapa de acondicionamiento y almacenamiento de la imagen.

El acondicionamiento de la imagen se realizó mediante un pulsador diseñado con el fin de que la imagen sea procesada cada vez que este sea presionado.

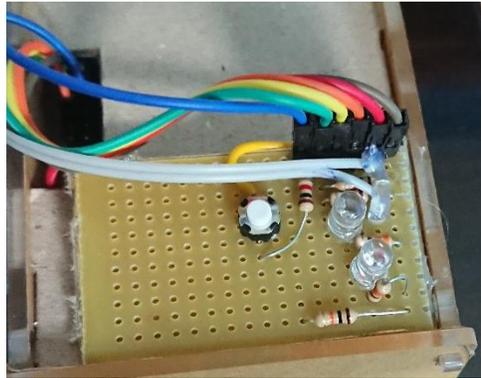


Figura 4.7. Circuito para la captura de imágenes – pulsador.

Fuente. – Elaborado por el Autor

Para el diseño del pulsador se utilizó un divisor de voltaje con resistencia pull down, de esta manera se garantiza un cambio de estado lógico ideal, como se muestra en la figura 4.7.

4.1.3 Etapa de diseño del algoritmo.

El algoritmo diseñado, para la conversión de palabras a audio, se basa en el uso de librerías de OpenCV compatibles con el lenguaje de programación Python, además tienen la característica de ser gratuitas y cumplen con los requerimientos de adquisición y procesamiento de imágenes, las cuales cumplen los siguientes procesos:

- Captura y almacenamiento de la imagen.
- Procesamiento de la imagen.
- Conversión a escala de grises.
- Segmentación.
- Extracción de caracteres y palabras.
- Conversión a audio.

4.1.4 Procesos realizados por el algoritmo.

En la figura 4.8, se muestra el algoritmo, para realizar los diferentes procesos necesarios para la implementación del proyecto.

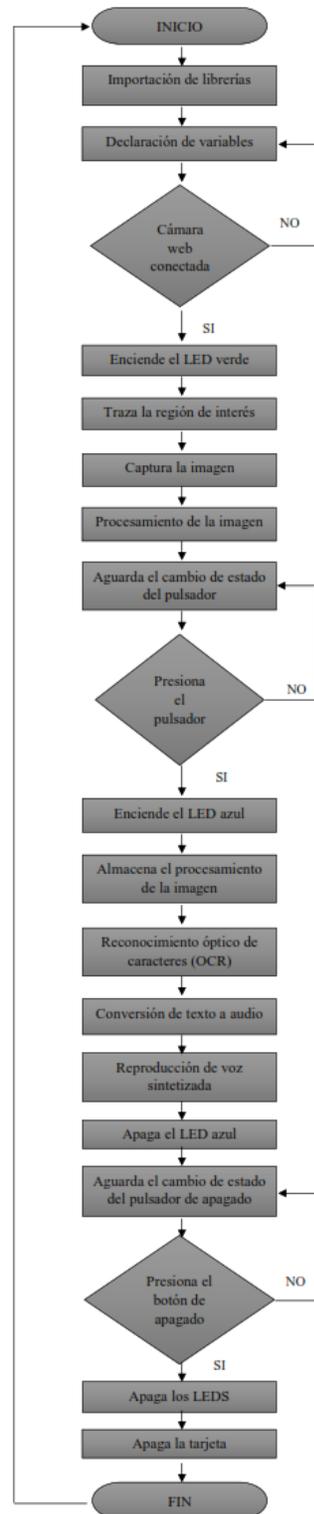


Figura 4.8. Diagrama de los procesos realizados por el algoritmo.

Fuente. – Elaborado por el Autor.

La figura 4.8, hace referencia a los diferentes métodos a realizar por el algoritmo, que consta de los siguientes procesos:

- Inicialmente el algoritmo realizara una importación de las librerías y se declara variables que serán necesarias para el tratamiento de la imagen.
- El algoritmo realiza una detección de sus puertos USB, en el caso de estar conectada la cámara se encenderá el led verde, seguidamente el algoritmo realiza el proceso de trazar la región de interés, luego realizará la captura de la imagen para efectuar un procesamiento de la misma que consiste en la conversión a escala de grises para finalmente limpiar dicha escala de grises mediante el proceso de umbral.
- Aguarda el cambio de estado lógico del pulsador, es decir uno o cero.
- Al presionar el pulsador el algoritmo procede a almacenar el proceso antes mencionado, para consecutivamente realizar el reconocimiento óptico de caracteres (OCR). Este procedimiento se ejecuta mientras el led azul se encuentra encendido.
- Tras efectuar el reconocimiento óptico de caracteres el algoritmo realiza la conversión del texto contenido en la imagen a audio por medio de la librería espeak.
- El siguiente proceso del algoritmo es reproducir en forma de voz sintetizada los caracteres antes reconocidos, la voz sintetizada se escuchará mediante un auricular.
- Finalmente, el algoritmo apaga el led azul y aguarda por el cambio de estado del botón de apagado, si el estado de este cambia, el algoritmo procede a apagar el del verde cerrar todas las ventanas virtuales abiertas para después apagar la tarjeta.

4.1.4.1 Etapa de procesamiento de la imagen.

Una vez que la imagen es capturada, es necesario efectuar un procesamiento de la misma, con el objetivo de modificar las condiciones de la imagen, obteniendo una nueva con mejor apariencia visual que permita al sistema de reconocimiento entregar mejores resultados.

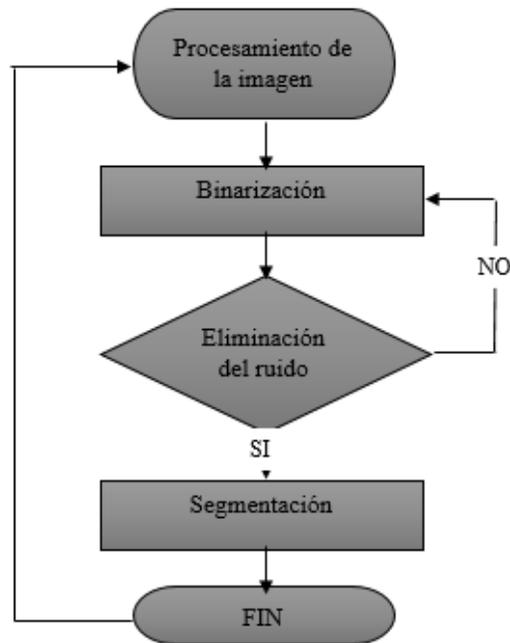


Figura 4.9. Diagrama del procesamiento de la imagen.

Fuente. – (Burga Caza, 2018).

En la figura 4.9, muestra el proceso que realiza el algoritmo que consiste en la captura de la imagen para una posterior conversión a escala de grises (binarización) de la misma con el fin de obtener una imagen en blanco y negro. Además, se procede a limpiar el proceso anterior por medio del método del umbral este método permite resaltar el color blanco y negro mediante la determinación de valores numéricos tanto para el color blanco y el color negro.

La función a utilizarse es `cv.threshold`, donde el primer argumento debe ser la imagen binarizada (convertida a escala de grises), el segundo argumento es el valor de umbral que se utilizara para clasificar los valores de pixeles. En OpenCV proporciona diferentes estilos de umbrales, en el capítulo de la propuesta se mencionó que se optara por el umbral binario siendo la función la siguiente.

- `cv.THRESH_BINARY`

Esta función limpia la escala de grises tomando en cuenta un valor umbral de conversión desde el color blanco hasta el color negro, el color blanco tiene un valor de 128, entonces para valores por debajo a este número, la función va resaltar el color blanco y para valores mayores a 255, va resaltar (pintar) de color negro, para devolver al final una imagen de blanco y negro,

consiguiendo así una limpieza adecuada al texto, de esta manera se conseguirá que el proceso de reconocimiento de caracteres realice una detección adecuada y más sencilla.

Finalmente se procede a realizar la segmentación proceso que consiste en extraer las regiones de interés de la imagen antes captada.

4.1.4.2 Etapa de reconocimiento óptico de caracteres (OCR).

En la figura 4.10, muestra el proceso que realiza el algoritmo, mismo que consiste en la examinación de la imagen anteriormente procesada para la realización del reconocimiento óptico de caracteres, a la vez el algoritmo realiza la comparación de los caracteres de la imagen, los mismos serán comparados con los caracteres pre existentes en la librería Python – Tesseract, que es parte del motor del OCR (reconocimiento óptico de caracteres).

Para una adecuada extracción de los caracteres se procedió a realizar algunas configuraciones al Tesseract. Lo primero que se configuro fue el idioma que por defecto es el inglés. Para configurar el idioma simplemente se especifica con *spa*. Otro aspecto que se configuró es el modo *psm 4* que indica, que la función Tesseract realizara una segmentación de la imagen y deberá analizar una cadena de texto en vertical con diferentes tamaños de letra, de esta manera se estará asegurando que el OCR sea capaz de reconocer diferentes tamaños de letra en el mismo párrafo. El último aspecto que se configuro fue *oem 3*, que es un modo de operación, basado en lo que está disponible, es decir que el OCR coloca la mejor palabra disponible que consiga en su librería, basándose en el reconocimiento de letras que forman la palabra, sin cambiar el contexto del texto.

Al no especificarse estas características el Tesseract entrara en un modo por defecto, caso que no es recomendable para el propósito del presente proyecto. El tiempo de procesamiento de la función Tesseract va depender de la complejidad de la imagen que se envié para su reconocimiento, ya que la función deberá descifrar y buscar letras o textos dentro de la imagen. Finalmente, el algoritmo guarda el texto reconocido para la realización de la lectura en voz sintetizada quedando de la siguiente manera:

- `pytesseract (thresh, config='-l spa --psm 4 --oem 3')`

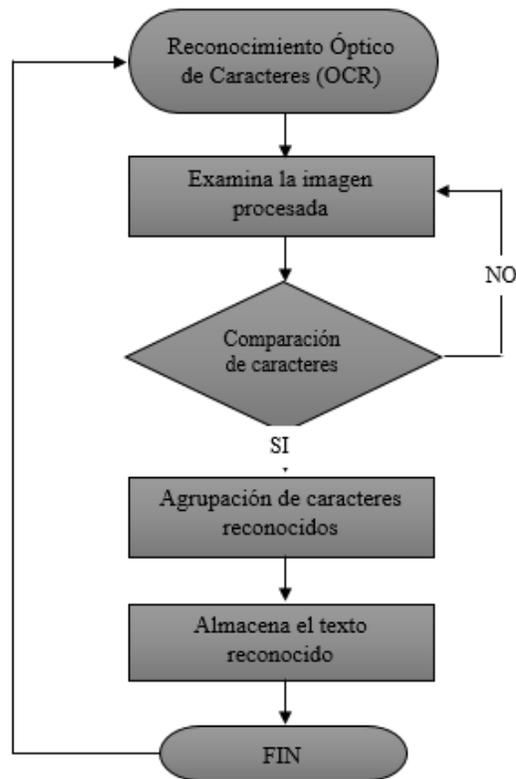


Figura 4.10. Diagrama del reconocimiento óptico de caracteres.

Fuente. – (Burga Caza, 2018).

4.1.4.3 Etapa de conversión a audio.

La figura 4.11, muestra la comparación de caracteres, el programa guardara el texto reconocido para la posterior lectura en voz sintetizada utilizando la librería espeak.

En este punto se realizó la configuración de la velocidad en que la voz sintetizada será reproducida, además del tipo de voz y la pausa entre palabras. Se decidió que:

- La voz será femenina.
- La velocidad de lectura será de 160 palabras por minuto (valor por defecto)
- La pausa entre palabras tendrá el valor por defecto.

Tomando en consideración el comando de línea quedaría de la siguiente manera:

- *'espeak', '-ves+f3'*

Para una adecuada lectura de la imagen procesada, la imagen capturada debe tener una buena calidad para que el algoritmo de reconocimiento óptico de caracteres (OCR) sea óptimo y se escuche el texto de una manera clara.

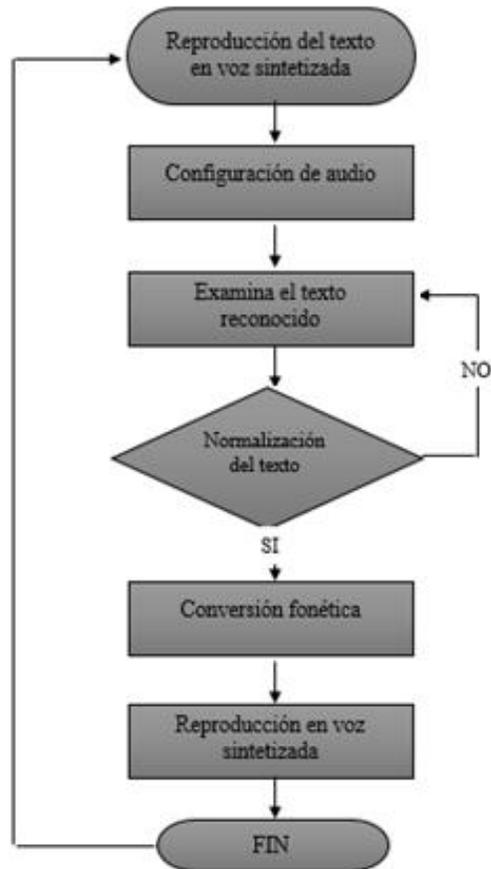


Figura 4.11. Diagrama de la reproducción del texto en voz sintetizada.

Fuente. – (Burga Caza, 2018).

4.2 Implementación.

La estructura final del dispositivo fue montada sobre una gorra y está compuesta por los siguientes elementos:

- La figura 4.12, muestra la fuente de alimentación, que está conformada por un power bank con un voltaje de 5 Voltios y una corriente de 2,1 Amperios.



Figura 4.12. Fuente de alimentación.

Fuente. - Elaborado por el autor.

- En la figura 4.13, se hace referencia a la carcasa que en su interior se encuentra la tarjeta Raspberry Pi y el circuito de control para el pulsador, que capturará la imagen para posterior procesamiento. La tarjeta Raspberry Pi permite interconectar los elementos adicionales como son: la cámara web, el pulsador y los auriculares.



Figura 4.13. Carcasa del sistema conectado.

Fuente. - Elaborado por el autor.

- Un auricular para la reproducción de la voz sintetizada, que se conecta directamente a la tarjeta Raspberry Pi, mediante el conector mini jack.
- Una cámara web para la captura de las imágenes, que posteriormente serán tratadas, la cámara se conecta directamente a la tarjeta mediante un puerto USB.

- Además, se implementó un sistema de iluminación adicional, que consta de una cinta led de forma cuadrada, la finalidad de este sistema es proveer de una iluminación frontal y lateral al documento a ser procesado. Dicho sistema funciona con un voltaje de 5 voltios que puede ser proporcionado o a su vez contar con una batería independiente. Como se muestra en la figura 4.14.

El sistema completo, se muestra en la figura 4.14, los elementos antes mencionados se montaron sobre una gorra para su uso por parte de la persona invidente.



Figura 4.14. Imagen del sistema convertidor de palabras a audio.

Fuente. - Elaborado por el autor.

4.3 Pruebas de funcionamiento.

4.3.1 Ambiente de prueba.

Para una adecuada captura de la imagen, el ambiente debe contar con las siguientes características:

- El documento a ser procesado debe contener palabras con letra número 18 y tipo Times New Roman.
- La distancia de la cámara hacia el documento debe ser de 17 centímetros.
- Las palabras contenidas en el documento deben estar en español.
- El usuario del prototipo, debe encontrarse en un lugar con una adecuada iluminación, además de contar con un documento que cumpla con las especificaciones antes señaladas.

- Una vez encendida la tarjeta, el usuario debe colocarse la gorra, que contiene todo el sistema y finalmente presionar el pulsador para que la imagen sea capturada.
- El usuario debe aguardar dependiendo de la cantidad de texto a ser procesado un tiempo que va desde los 5 segundos a los 20 segundos aproximadamente, para poder escuchar el contenido de la imagen mediante la voz sintetizada.
- El algoritmo está diseñado para realizar lecturas continuas, con la disposición de un botón que sirve para apagar el sistema completo.

4.3.2 Pruebas del prototipo

Una vez implementado el prototipo, se procedió a realizar las pruebas que aseguren la funcionalidad y permitan determinar resultados que valoren el trabajo invertido.

- Como muestra del funcionamiento del sistema, se procedió a realizar la lectura del siguiente documento. La figura 4.15, muestra los elementos que componen el prototipo, que está compuesto de un soporte donde se coloca los documentos y la gorra que contiene la tarjeta y los elementos periféricos.

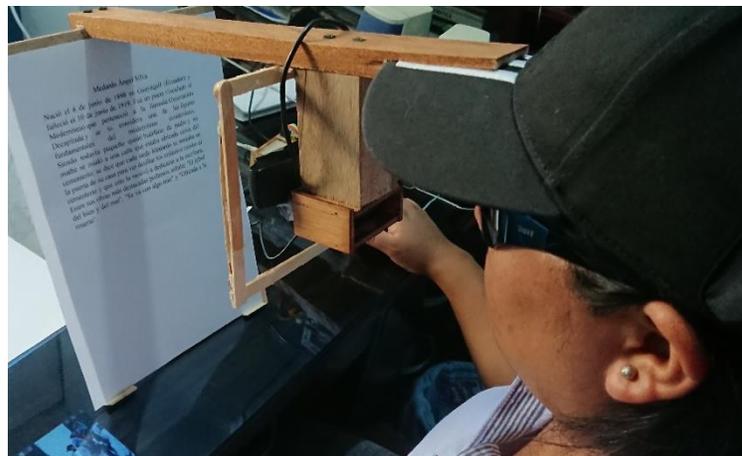


Figura 4.15 Dispositivo implementado.

Fuente. Elaborado por el autor.

Luego de que el prototipo realice la detección de la región de interés de la imagen, esta pasa por algunos procesos entre los que destacan:

- Captura de la región de interés de la imagen; Figura 4.16.

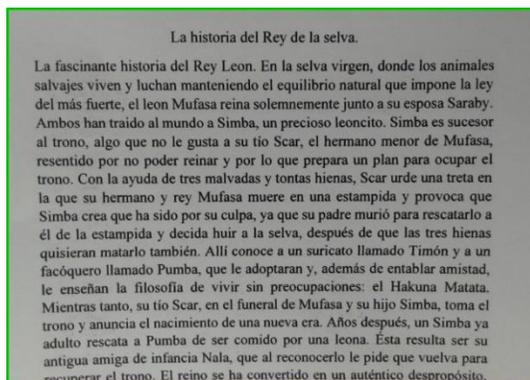


Figura 4.16. Captura de la región de interés de la imagen.

Fuente. – Elaborado por el autor.

- Conversión a escala de grises; Figura 4.17.

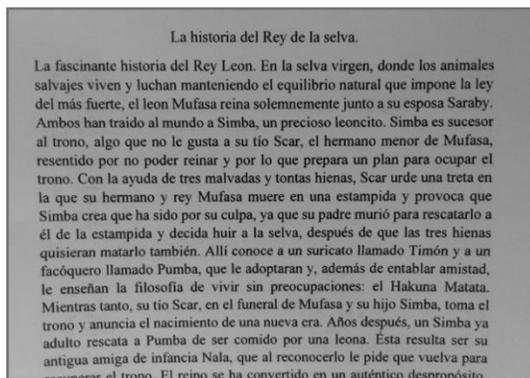


Figura 4.17. Conversión a escala de grises.

Fuente. – Elaborado por el autor.

- Extracción de caracteres. Figura 4.18

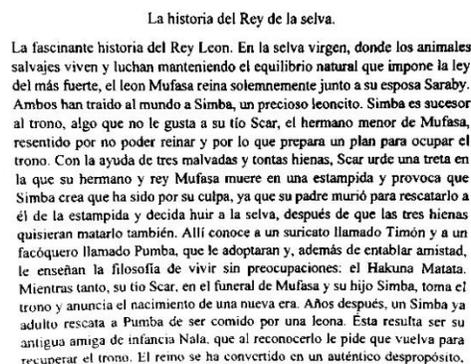


Figura 4.18. Extracción de caracteres.

Fuente. – Elaborado por el autor.

Como parte final del proceso se escuchará la voz sintetizada por medio de los auriculares, del texto que el algoritmo reconoció. Siendo esta la última etapa del algoritmo diseñado.

4.4 Análisis de resultados.

En el apartado anterior se muestra las diferentes pruebas realizadas a textos con diferente contenido y con un tamaño de letra en el rango de 12 hasta 18 con la posibilidad de realizar una correcta lectura de textos de tamaño de letra superior a 18, como se muestra en la tabla 4.1.

Tabla 4.1. Documentos analizados.

Documentos Analizados		
Tamaño de letra	Lectura Correcta	Lectura Incorrecta
Número 12		X
Número 14	X	
Número 16	X	
Número 18	X	
Número 20	X	

Fuente. Elaborado por el autor.

Se comprobó entonces que existe una correcta lectura para textos con letras de tamaño que están en el rango de 14 a 18 y superiores. Cabe mencionar que dichas pruebas se realizaron a una distancia de 17 centímetros desde el lente de la cámara hacia el texto a ser leído. También se observó que, para documentos con textos de tamaño de letra inferior a 12, la distancia de 17 centímetros influye directamente en la correcta detección del texto contenido, que para este caso mostro inconsistencias y errores en las palabras por parte del OCR y a la vez que el tiempo de procesamiento fue mayor.

Finalmente, la carencia de una adecuada iluminación también afecta al correcto funcionamiento del prototipo, ya que, al capturar una imagen con una luz tenue, el OCR confunde las letras y presenta símbolos en lugar de letras debido a, que no puede hacer una comparación adecuada. No obstante se pudo comprobar que al existir una buena iluminación no fue necesario emplear el dispositivo de iluminación adicional con el que cuenta el sistema.

CONCLUSIONES

- El tema planteado consistió en reconocer palabras usando una tarjeta Raspberry Pi mediante una cámara web, por lo que, una vez que se llevó a cabo la metodología se obtuvieron resultados que muestran la posibilidad de reconocer palabras contenidas en una imagen, por lo tanto, se concluye que el tema puede ser resuelto si se implementa la propuesta de solución.
- Se encontró que los parámetros de funcionamiento de la tarjeta Raspberry Pi, fueron comprendidos y empleados exitosamente en el desarrollo del algoritmo a ser implementado para el funcionamiento del prototipo.
- Se comprobó que el algoritmo desarrollado para el reconocimiento de las palabras contenidas en la imagen, emplea programas libres que cuentan con una precisión adecuada, que pueden ser utilizados para el diseño de prototipos enfocados a la ayuda social.
- Se determinó que, tras la realización de las pruebas de funcionamiento del prototipo por parte de la persona invidente, la misma considera que, el dispositivo permite la conversión de palabras o texto a audio.
- Se observó que, para documentos con el tipo de letra Times New Roman y tamaño número 12, a una distancia superior a los 17 centímetros del documento al lente de la cámara, el algoritmo no fue capaz de realizar una correcta detección del texto contenido por la imagen, confundiendo letras con símbolos.
- Se comprobó que, para documentos con el tipo de letra Times New Roman y para tamaños que están en el rango de 14 a 18 y superiores, el algoritmo diseñado es capaz de realizar una correcta detección de texto contenido por la imagen.
- Se comprobó que el algoritmo diseñado es capaz de realizar una detección correcta para documentos con un tipo de letra diferente al tipo de letra propuesto.

- Se comprobó que, la falta de una buena iluminación afecta al correcto funcionamiento del prototipo, ya que, al capturar una imagen con una luz tenue, el OCR confunde las letras y presenta símbolos en lugar de letras.

RECOMENDACIONES

- Se recomienda, en primer lugar, emplear una cámara que cuente con una mejor resolución en el ámbito de densidad de píxeles, además de contar con sensores de iluminación, de esta manera se asegura la captura de una imagen más nítida y con mejor iluminación. (auto enfoque y auto zoom)
- Asimismo, se recomienda que la distancia promedio para una adecuada captura de la imagen la distancia no debe exceder los 17 cm, ya que este valor asegura una captura completa del documento a ser reconocido por el algoritmo.
- Además, se debe implementar un sistema de enfriamiento, ya que el uso continuo de la tarjeta Raspberry Pi, por un lapso de dos horas, los elementos y procesador tienen a calentarse.
- Es importante que el prototipo no sufra duras caídas, para evitar el daño de uno de los elementos periféricos, además que el prototipo no debe exponerse a climas muy cálidos o lluviosos.

REFERENCIAS BIBLIOGRÁFICAS

- Arevalo, V., González, J., & Ambrosio, G. (2016). *La librería de visión artificial OpenCV aplicación a la docencia e investigación*. Málaga.
- AutoSolar. (19 de 04 de 2015). Obtenido de <https://autosolar.es/blog/aspectos-tecnicos/como-calcular-la-capacidad-de-las-baterias>
- Ayala, E., & Vásquez, F. (2012). *Prototipo de un sistema de ultrasonido aplicado a un*. Quito.
- Barrobés, H., & Ruiz, M. (2017). *Síntesis del habla*. Catalunya.
- DCM SISTEMES. (11 de Agosto de 2018). *Documentación detallada y descargas*. Obtenido de <https://www.dcmsistemas.com/es/documentacion/informacion-adicional/44/el-color-de-la-luz/>
- Esqueda Elizondo, J. J. (2002). *Fundamentos de Procesamiento de Imágenes*. Mexico.
- García, A. (20 de 10 de 2018). *DIGILOGIC*. Obtenido de <http://www.digilogic.es/que-son-las-resistencias-pull-up/>
- Hurtado, J. (21 de 02 de 2008). *La Investigación Proyectiva*. Obtenido de <http://investigacionholistica.blogspot.com/2008/02/la-investigacin-proyectiva.html>
- Latin Fotografía*. (25 de 9 de 2018). Obtenido de <http://www.latinfotografia.net/foros/index.php?topic=39361.0>
- Llamas, L. (23 de 08 de 2018). *Ingeniería, informática y diseño*. Obtenido de <https://www.luisllamas.es/raspberry-pi-instalar-noobs/>
- logitech. (25 de 11 de 2018). *Logitech*. Obtenido de <https://www.logitech.com/es-es/product/c922-pro-stream-webcam>

- Murillo Quishpe , G. E., & Montaluisa Pilatasig, S. D. (2010). *Control e inspeccion de llenado de botellas aplicando herramientas de vision artificial para el laboratorio de neumatica e hidronica de la espe sede Latacunga*. Latacunga.
- Nogue, A., & Antiga, J. (2012). Aplicacion practica de la vision artificial en el control de procesos industriales. *Conocimientos basicos de Vision Artificial*, 9-15.
- Ñiacasha, N. (2004). *Desarrollo de un dispositivo que mida la distancia a un objeto*. Quito.
- Ortuño López, J. J. (2016). *Detección e identificacion visual de caracteres en productos industriales*. Alicante.
- Palacios Sanchez, R. (2018). *Reconocimientos de imagenes mediante Raspberry Pi*. España.
- PROMETEC. (11 de 12 de 2018). Obtenido de <https://www.prometec.net/>
- Raspberry Shop. (10 de Diciembre de 2018). Obtenido de <https://www.raspberrystore.com/index.php>
- Sierra Álvarez, S. (2012). *Sistema de medicion de objetos basado en vision artificial*. Medellín.
- SOURCEFORGE. (5 de 11 de 2018). Obtenido de <http://espeak.sourceforge.net/>

ANEXOS

MANUAL PARA EL USUARIO

**DISPOSITIVO QUE MEDIANTE VISIÓN ARTIFICIAL
PERMITA ADQUIRIR IMÁGENES CON PALABRAS PARA LA
CONVERSIÓN A AUDIO, ORIENTADO A LA AYUDA DE
PERSONAS INVIDENTES.**



ÍNDICE

ÍNDICE.....	69
INTRODUCCIÓN.....	70
1. MÓDULO DE CONTROL.....	70
2. MODULO DE ALIMENTACIÓN.....	70
2.1. CARGADOR DEL POWER BANK.....	71
2.2. RECOMENDACIONES PARA LA CARGA DEL POWER BANK.....	71
2.3. PROCESO DE CARGA.....	72
3. PUESTA EN MARCHA.....	72
3.1. ENCENDIDO DEL SISTEMA:.....	72
3.2. COLOCACIÓN DEL DISPOSITIVO:.....	72
3.3. CAPTURA DE LA IMAGEN:.....	73
3.4. DOCUMENTO A SER PROCESADO:.....	74
3.5. REPRODUCCIÓN DEL TEXTO PROCESADO:.....	74
4. APAGADO DEL SISTEMA.....	74
5. RECOMENDACIONES.....	74

INTRODUCCIÓN.

El presente manual, contiene información sobre el funcionamiento básico del dispositivo conversor de palabras a audio. Con la finalidad de realizar una adecuada manipulación por parte de la persona de apoyo.

El dispositivo está compuesto por dos módulos: el módulo de control y el módulo de alimentación.

1. MÓDULO DE CONTROL.

Compuesto por una carcasa que contiene a la tarjeta Raspberry Pi, el circuito de control del pulsador y los dispositivos periféricos.

La tarjeta Raspberry Pi, que es la encargada de realizar todo el proceso de tratamiento de la imagen además de interconectar los elementos adicionales como son: la cámara, el pulsador y los auriculares. Los elementos adicionales se conectan directamente a la tarjeta mediante USB (cámara), pines GPIO (pulsador) y leds



Figura 1. Módulo de control del dispositivo.

2. MODULO DE ALIMENTACIÓN

Básicamente se compone de un power bank, que brinda energía tanto a la tarjeta Raspberry Pi y a la vez al resto de componentes del sistema. Para un uso constante del dispositivo el power bank tiene la característica de ser recargable, conectándola mediante un cargador directo a la toma de la pared.



Figura 2. Módulo de alimentación del dispositivo.

Las características del power bank usado para la alimentación del sistema son las siguientes:

- Voltaje de salida: 5 Voltios.
- Corriente de salida: 2.1 Amperios.

Adicionalmente el módulo de alimentación tiene la característica de ser autoencendido una vez conectado a la tarjeta, de igual manera, el circuito del pulsador cuenta con un botón de pagado para de esta manera evitar el desgaste del power bank.

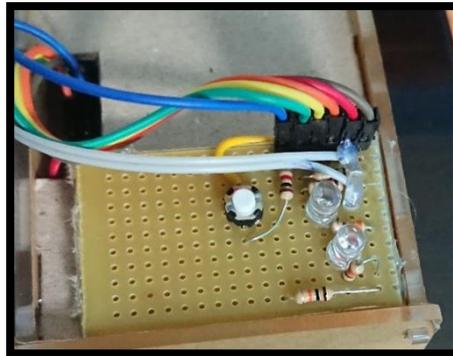


Figura 3. Carcasa del módulo de control del dispositivo.

2.1. CARGADOR DEL POWER BANK

El proceso de carga del power bank es muy sencillo, se puede realizar la carga por medio de una entrada micro USB. Se recomienda utilizar un cargador de celular, que permite extender la vida útil de la batería interna del power bank.

2.2. RECOMENDACIONES PARA LA CARGA DEL POWER BANK.

- Verificar que la conexión sea segura. (interiores)
- Verificar la correcta polaridad de la batería y dispositivo de carga.
- Verificar que el método de carga sea el correcto.

2.3. PROCESO DE CARGA.

Durante el proceso de carga es importante mantener la atención a los cuatro leds que se encuentran sobre la parte superior que permite verificar el estado de la batería del power bank. Los leds según el nivel de batería indican:

- 1 LED: 0%-25%
- 2 LEDs: 25%-50%
- 3 LEDs: 50%-75%
- 4 LEDs: 75%-100%

3. PUESTA EN MARCHA.

Para un adecuado funcionamiento del sistema, es necesario seguir las siguientes instrucciones:

3.1. ENCENDIDO DEL SISTEMA:

Para encender el sistema únicamente basta con conectar la tarjeta por medio del cable micro USB y esta se encenderá automáticamente. Se recomienda aguardar por un periodo de 1 minuto como máximo hasta que el sistema operativo cargue por completo.

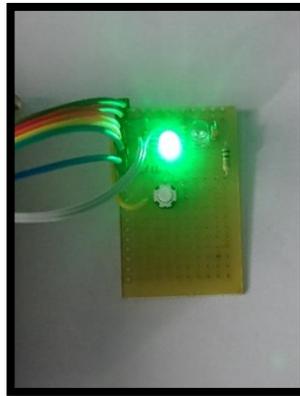


Figura 4. Led de encendido del sistema.

Una vez que el sistema operativo termine de cargar, se encenderá un led de color verde, esto indica que el algoritmo está listo para funcionar.

3.2. COLOCACIÓN DEL DISPOSITIVO:

Al tratarse de un dispositivo montado sobre una gorra. El dispositivo se colocará en la cabeza del usuario con la visera hacia el frente, ya que la cámara que capturará la imagen se encuentra ubicada bajo un soporte que esta sujeta a la visera de la gorra.

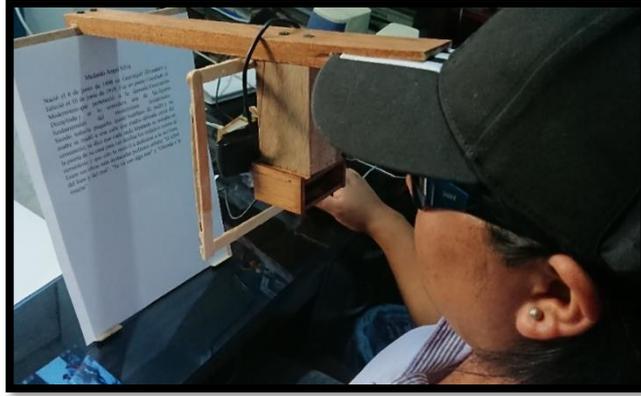


Figura 5. Colocación del dispositivo.

3.3. CAPTURA DE LA IMAGEN:

Luego de colocarse la gorra sobre la cabeza, se procederá a capturar la imagen a ser procesada. Mediante el pulsador que se encuentra al lado izquierdo de la gorra. Únicamente se necesita un pulso para que la imagen sea capturada para posterior ser procesada.

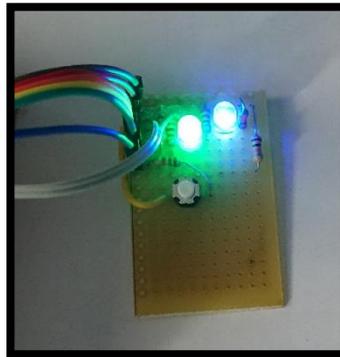


Figura 6. Captura de la imagen.

Una vez presionado el pulsador, se encenderá un led azul, este indica que el algoritmo diseñado esta realizado el proceso de tratamiento de la imagen. Una vez se termine de escuchar la voz sintetizada, este se apagará, indicando que se puede proceder a la lectura de otro documento.

Nota: Se recomienda que la visera de la gorra este alineada con el documento de interés para la posterior captura de la imagen. Esta distancia no debe ser mayor a 18 cm, dicha distancia asegura un adecuado reconocimiento óptico de caracteres por parte del algoritmo empleado por el sistema.

3.4. DOCUMENTO A SER PROCESADO:

Se recomienda que el documento a ser capturado en forma de imagen debe contener las siguientes características: el tamaño del documento que contiene el texto a ser procesado debe ser 21 x 29,7 cm y un tamaño de letra número 18. No obstante el dispositivo cuenta con la posibilidad de reconocer textos de documentos con letras más pequeñas.

3.5. REPRODUCCIÓN DEL TEXTO PROCESADO:

El audio convertido en voz sintetizada, se lo realizara mediante un auricular que se colocaran directamente en los oídos del usuario.

Se recomienda que el uso del dispositivo no debe exceder las dos horas consecutivas.

4. APAGADO DEL SISTEMA.

Para apagar el sistema completo simplemente se debe pulsar el botón que se encuentra en el módulo de control del sistema.

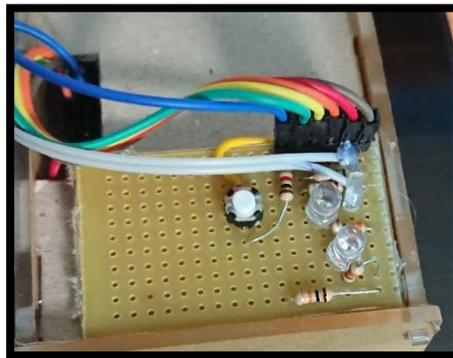


Figura 7. Botón de apagado del sistema.

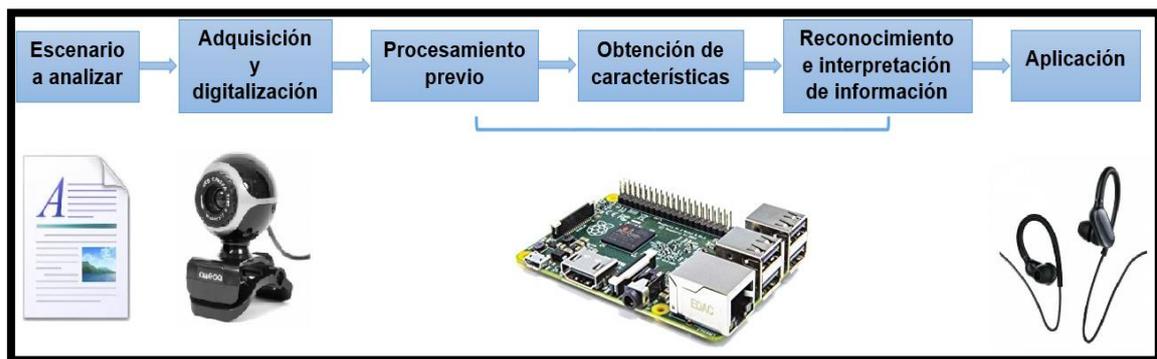
Una vez presionado el botón, el led verde se apagará y la tarjeta procederá a realizar un apagado completo.

5. RECOMENDACIONES

- La captura de la imagen debe realizarse en un lugar con una iluminación adecuada.
- Realice una carga completa del power bank, de esta manera se asegura un adecuado funcionamiento del sistema.
- Mantenga los cables de los módulos de control y de alimentación sin nudos ni acodamientos.
- Luego del uso del sistema se recomienda realizar la recarga de la batería por un lapso de 2 horas como máximo.
- Revise el estado de los auriculares ya que el deterioro de los mismos, impedirá el adecuado funcionamiento del sistema.

MANUAL TÉCNICO

**DISPOSITIVO QUE MEDIANTE VISIÓN ARTIFICIAL
PERMITA ADQUIRIR IMÁGENES CON PALABRAS PARA LA
CONVERSIÓN A AUDIO, ORIENTADO A LA AYUDA DE
PERSONAS INVIDENTES.**



ÍNDICE

ÍNDICE.....	76
INTRODUCCIÓN.....	77
OBJETIVO.....	77
1. ESPECIFICACIONES TÉCNICAS.....	77
2. CONFIGURACIÓN DEL PROTOTIPO.....	77
2.1. CARACTERÍSTICAS TÉCNICAS DE LA TARJETA.....	77
2.2. INSTALACIÓN DEL SISTEMA OPERATIVO.....	78
2.3. CONFIGURACIÓN DEL ALGORITMO IMPLEMENTADO.....	81
3. CASOS DE USO.....	83
3.1. AMBIENTE DE PRUEBA.....	83
4. RECOMENDACIONES.....	83

INTRODUCCIÓN.

La finalidad del presente manual técnico es la de proporcionar al lector la lógica con la que se ha desarrollado una aplicación, la cual se sabe que es propia del autor; por lo que se considera necesario ser documentada.

Se aclara que el manual no pretende ser un curso de aprendizaje de cada una de las herramientas empleadas para el desarrollo del sistema. Para un mayor detalle y acerca de cada una de las herramientas utilizadas, y su forma de operación y aplicación, se recomienda consultar los manuales respectivos de cada una de ellos.

OBJETIVO.

Proporcionar una guía para el lector, del desarrollo del prototipo y del correcto funcionamiento de los diferentes elementos que lo componen. Y de la instalación de los elementos periféricos.

1. ESPECIFICACIONES TÉCNICAS.

El prototipo desarrollado, está compuesto principalmente por una tarjeta Raspberry Pi modelo 3 B. Que es la encargada de realizar los diferentes procesos del tratamiento de imágenes, cuyas características físicas se muestran en la figura 1.

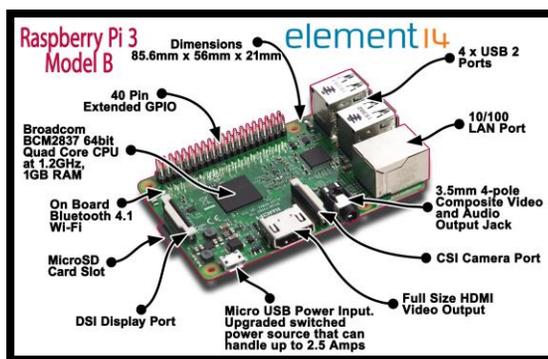


Figura 19. Características físicas de la Tarjeta Raspberry Pi 3B.

2. CONFIGURACIÓN DEL PROTOTIPO.

2.1. CARACTERÍSTICAS TÉCNICAS DE LA TARJETA.

En la tabla 1, se muestra las características técnicas de la tarjeta Raspberry Pi 3.

COMPONENTE	DESCRIPCIÓN
GPIO I/O 40 pines	Pines de entrada y salida digitales, 26 son pines GPIO y los otros son pines de alimentación.
Microprocesador Broadcom BCM2837	Consta cuatro núcleos de procesamiento ARM Cortex-A53 de alto rendimiento que funcionan a 1.2 GHz
Bluetooth 4.1, Wifi	Tecnologías de comunicación inalámbricas
Conexión display	Sirve para conectar una pantalla táctil.
Ranura Micro SD	Sirve para el almacenamiento de datos, cargar el sistema operativo en la tarjeta Raspberry Pi 3.
Ranura Micro USB	Alimentación 5 V- hasta 2,4 A
Salida multimedia HDMI	Conexión de video, audio
Conexión cámara MIPI CSI-2	Facilita la conexión de una cámara pequeña al procesador
Salida audio jack 3,5mm	Permite la conexión de parlantes o auriculares.
Puerto LAN 10/100	RJ 45 para la conexión a internet.
USB 2,0	4 puertos USB para la conexión de dispositivos.

Tabla 2. Características técnicas de la tarjeta Raspberry Pi.

2.2. INSTALACIÓN DEL SISTEMA OPERATIVO.

El prototipo fue desarrollado basado en el sistema operativo Raspbian. Para un adecuado proceso de instalación se procederá de la siguiente manera:

1.- Descargar NOOBS: En primer lugar, se descargó Noobs desde la página web de Raspberry Pi. Cabe mencionar que se debe descargar la versión completa ya que la misma contiene el software Raspbian, que es el sistema operativo que se instaló, como se muestra en la figura 2.

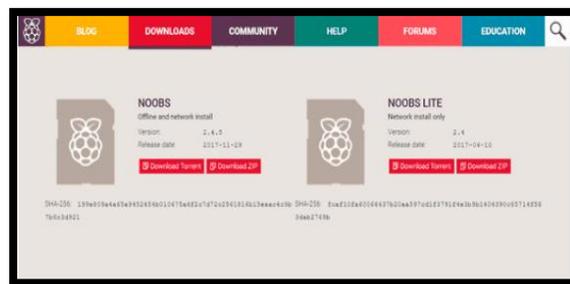


Figura 20. Gestor de instalación de sistemas operativos Noobs.

2.- Formatear la tarjeta microSD: En la figura 3, se muestra el proceso de formateo de la tarjeta micro SD, se recomienda utilizar una herramienta para formatear tarjetas microSD, para este caso se utilizó SD Card Formatter que está disponible de forma gratuita. Es aconsejable realizar el formateo de la tarjeta microSD para no tener errores de software en el futuro.

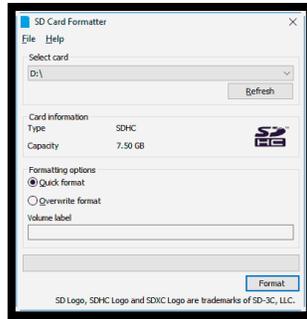


Figura 21. Cuadro de dialogo SD Card Formatter.

3.- Instalar Noobs: Se descomprime el fichero antes descargado, una vez que el fichero termine de copiarse en su totalidad en la tarjeta microSD. Se procedió a insertar en la tarjeta Raspberry Pi.

4.- Instalar Raspbian: Una vez conectada la tarjeta Raspberry Pi, a la toma de corriente, al iniciar aparecerá un menú con los distintos sistemas operativos disponibles, se eligió el sistema operativo Raspbian, como se muestra en la figura 4.

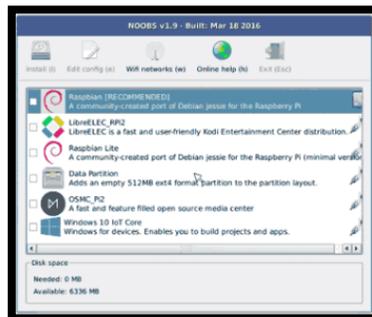


Figura 22. Elección del sistema operativo Raspbian.

La figura 5, muestra el proceso de instalación del sistema operativo Raspbian, en la tarjeta de memoria micro SD.

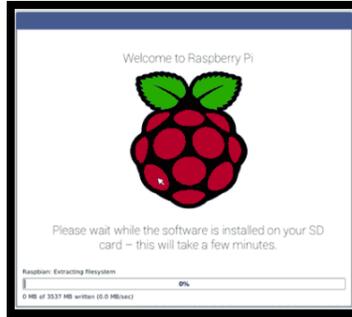


Figura 23. Proceso de instalación del sistema operativo Raspbian.

Una vez que el sistema operativo termina de instalarse por completo, es necesario reiniciar la tarjeta Raspberry Pi, para que el sistema operativo cargue, como se muestra en la figura 6.

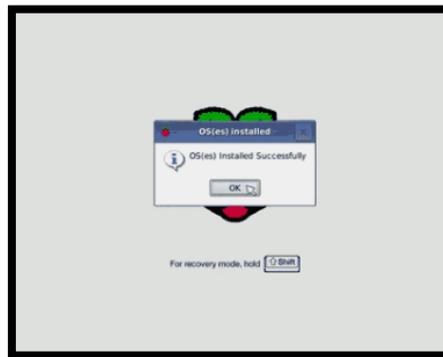


Figura 24. Finalización de la instalación del sistema operativo Raspbian.

6.- Reinicio de la tarjeta: Finalmente, la tarjeta Raspberry Pi se reinicia y ejecuta el sistema operativo instalado, siendo esta la parte final del proceso de instalación, como muestra la figura 7.



Figura 25. Pantalla de inicio del sistema operativo Raspbian.

2.3. CONFIGURACIÓN DEL ALGORITMO IMPLEMENTADO

Básicamente el algoritmo se compone de librerías propias de OpenCV, compatibles con el lenguaje de programación Python, además tienen la característica de ser gratuitas y cumplen con los requerimientos de adquisición y procesamiento de imágenes, las cuales cumplen los siguientes procesos:

- Captura y almacenamiento de la imagen.
- Procesamiento de la imagen.
- Conversión a escala de grises.
- Segmentación o conversión a blanco y negro de la imagen.
- Extracción de caracteres y palabras.
- Conversión a audio.

La figura 8, muestra los procesos realizados por el algoritmo, implementados para el tratamiento y posterior reconocimiento del contenido de la imagen capturada.

- Inicialmente el algoritmo realiza una importación de las librerías y se declara variables que serán necesarias para el tratamiento de la imagen.
- El algoritmo realiza una detección de sus puertos USB, en el caso de estar conectada la cámara se encenderá el led verde, esto indica que el algoritmo realiza el proceso de trazar la región de interés, seguidamente realizará la captura de la imagen para efectuar un procesamiento de la imagen que consiste en la conversión a escala de grises para finalmente limpiar dicha escala de grises mediante el proceso de umbral.
- Aguarda el cambio de estado lógico del pulsador, es decir uno o cero.
- Al presionar el pulsador el algoritmo procede a almacenar el proceso antes mencionado, para consecutivamente realizar el reconocimiento óptico de caracteres (OCR). Este procedimiento se ejecuta mientras el led azul se encuentra encendido.
- Tras efectuar el reconocimiento óptico de caracteres el algoritmo realiza la conversión del texto contenido en la imagen a audio por medio de la librería espeak.
- El siguiente proceso del algoritmo es reproducir en forma de voz sintetizada de los caracteres antes reconocidos, la voz sintetizada se escuchará mediante un auricular.
- Finalmente, el algoritmo apaga el led azul y aguarda por el cambio de estado del botón de apagado, si el estado de este cambia, el algoritmo procede a apagar el led verde cerrar todas las ventanas virtuales abiertas para después apagar la tarjeta.

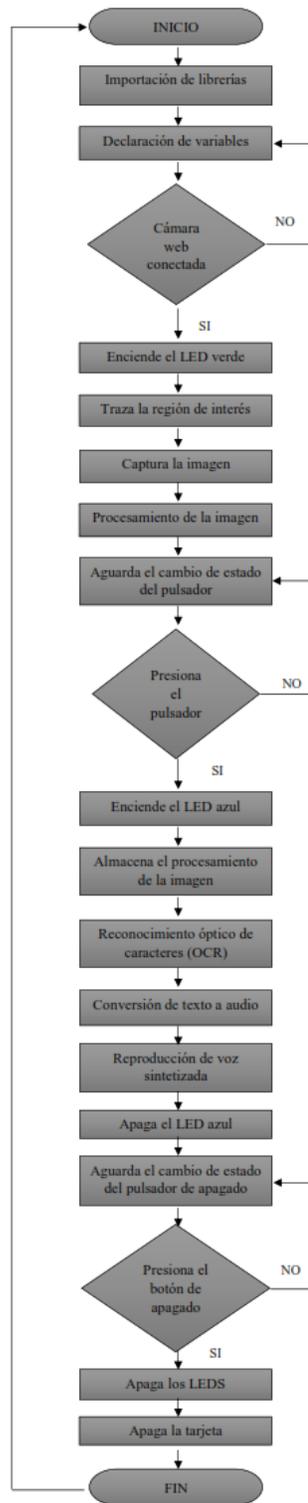


Figura 26. Diagrama de los procesos realizados por el algoritmo implementado.

3. CASOS DE USO.

El propósito del desarrollo del prototipo es la conversión de imágenes con palabras para convertirlas a audio. Es decir que el principal usuario, es una persona invidente.

A continuación, se detallan algunas características, que se deben tener en cuenta para su correcto funcionamiento.

3.1. AMBIENTE DE PRUEBA.

Para una adecuada captura de la imagen, el ambiente debe contar con las siguientes características:

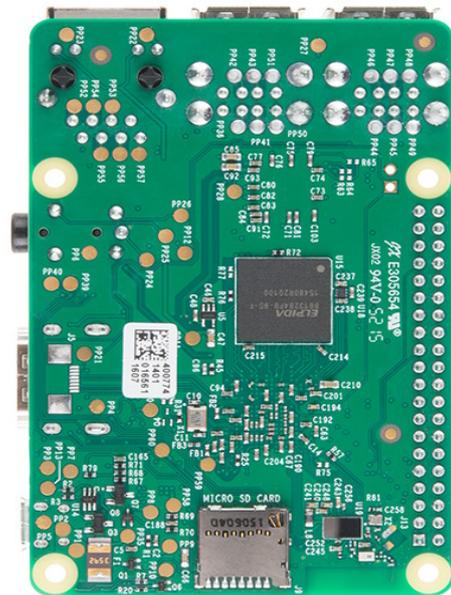
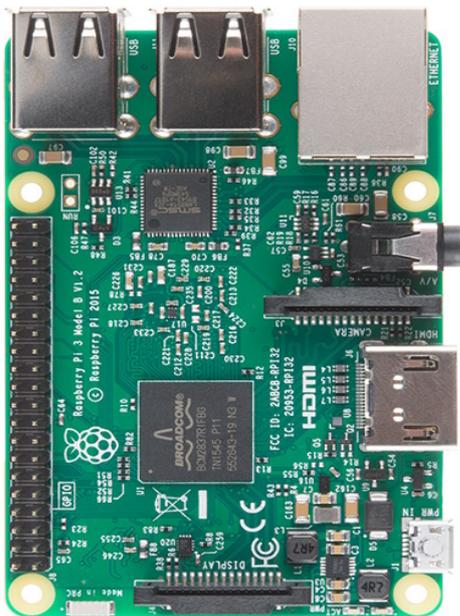
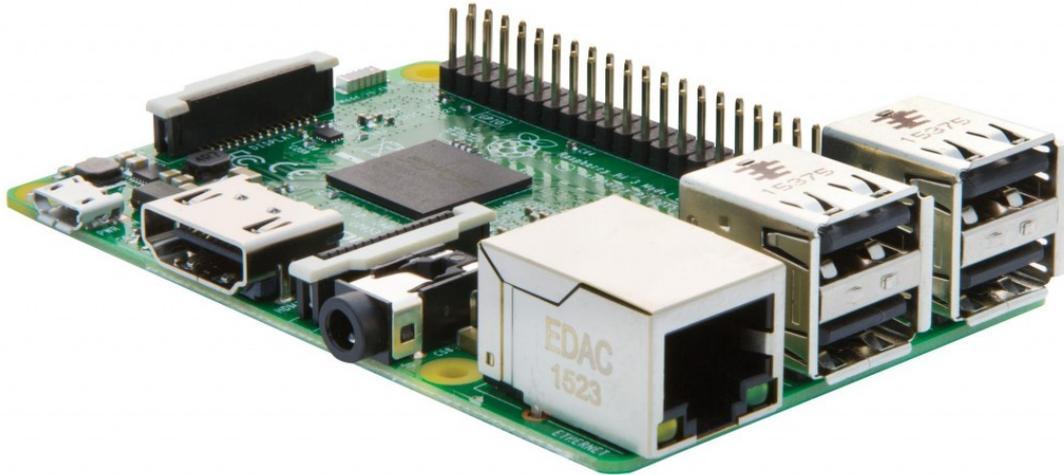
- El usuario del prototipo, debe encontrarse en un lugar con una adecuada iluminación, además de contar con un documento que cumpla con las especificaciones antes señaladas.
- La distancia de la cámara hacia el documento no debe exceder de los 18 centímetros.
- El usuario del prototipo, debe encontrarse en un lugar con una adecuada iluminación, además de contar con un documento que cumpla con las especificaciones antes señaladas.
- Una vez encendida la tarjeta, el usuario debe colocarse la gorra, que contiene todo el sistema y finalmente presionar el pulsador para que la imagen sea capturada.
- El usuario debe aguardar dependiendo de la cantidad de texto a ser procesado un tiempo que va desde los 5 segundos a los 20 segundos aproximadamente, para poder escuchar el contenido de la imagen mediante la voz sintetizada.
- El algoritmo está diseñado para realizar lecturas continuas, con la disposición de un botón que sirve para apagar el sistema completo.

4. RECOMENDACIONES

- En el caso de ser necesario un reinicio del prototipo, se debe proceder a apagar la tarjeta por medio del botón ubicado en el módulo de control del sistema.
- Revisar el estado de carga de la batería, ya que la descarga de la misma afecta directamente el correcto funcionamiento del prototipo.
- Revisar que los elementos periféricos (cámara, auricular, pulsador, leds) que componen el prototipo, estén conectados de forma correcta y los cables no presenten nudos.
- De presentarse alguna duda con respecto al funcionamiento del prototipo, recurrir al manual del usuario.

TARJETA RASPBERRY PI

RASPBERRY PI 3 MODEL B



Product Name: RASPBERRYPI3-MODB-1GB

Technical Specification:

Processor

- Broadcom BCM2387 chipset.
- 1.2GHz Quad-Core ARM Cortex-A53 (64Bit)

802.11 b/g/n Wireless LAN and Bluetooth 4.1 (Bluetooth Classic and LE)

- IEEE 802.11 b / g / n Wi-Fi. Protocol: WEP, WPA WPA2, algorithms AES-CCMP (maximum key length of 256 bits), the maximum range of 100 meters.
- IEEE 802.15 Bluetooth, symmetric encryption algorithm Advanced Encryption Standard (AES) with 128-bit key, the maximum range of 50 meters.

GPU

- Dual Core Video Core IV® Multimedia Co-Processor. Provides Open GL ES 2.0, hardware-accelerated Open VG, and 1080p30 H.264 high-profile decode.
- Capable of 1Gpixel/s, 1.5Gtexel/s or 24GFLOPs with texture filtering and DMA infrastructure

Memory

- 1GB LPDDR2

Operating System

- Boots from Micro SD card, running a version of the Linux operating system or Windows 10 IoT

Dimensions

- 85 x 56 x 17mm

Power

- Micro USB socket 5V1, 2.5A

Connectors:

Ethernet

- 10/100 BaseT Ethernet socket

Video Output

- HDMI (rev 1.3 & 1.4)
- Composite RCA (PAL and NTSC)

Audio Output

- Audio Output 3.5mm jack
- HDMI
- USB 4 x USB 2.0 Connector

GPIO Connector

- 40-pin 2.54 mm (100 mil) expansion header: 2x20 strip
- Providing 27 GPIO pins as well as +3.3 V, +5 V and GND supply lines

Camera Connector

- 15-pin MIPI Camera Serial Interface (CSI-2)

Display Connector

- Display Serial Interface (DSI) 15 way flat flex cable connector with two data lanes and a clock lane

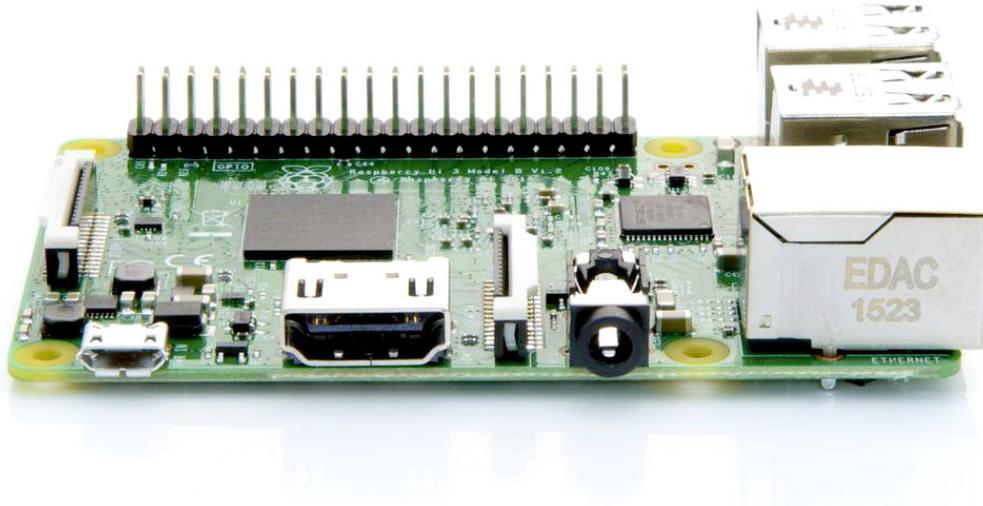
Memory Card Slot

- Push/pull Micro SDIO

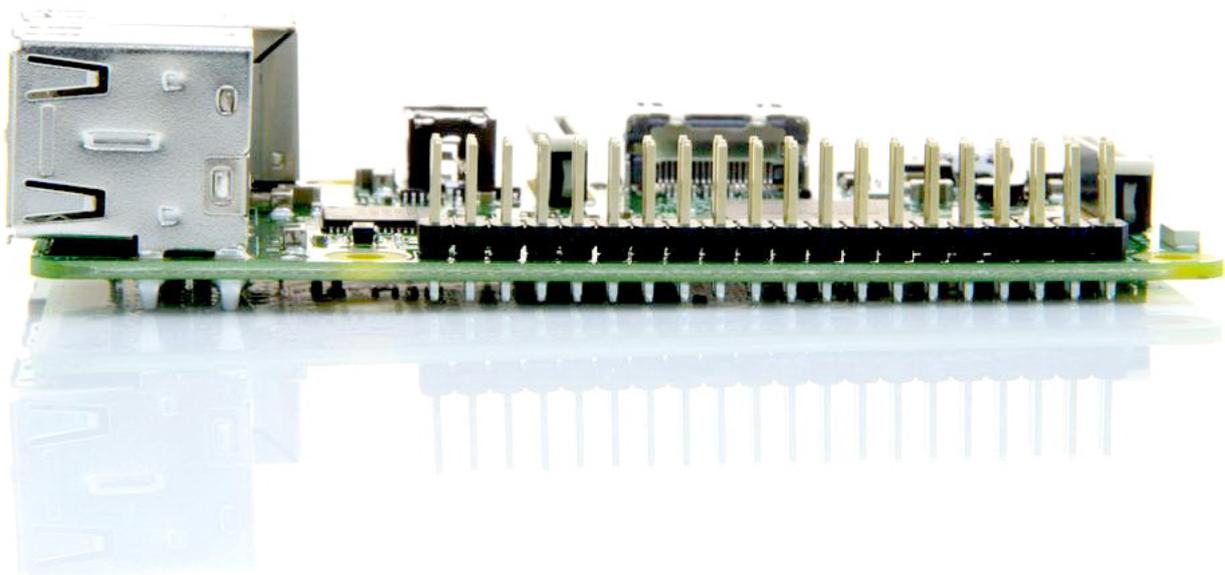
<http://uk.farnell.com/buy-raspberry-pi>

<http://www.newark.com/buy-raspberry-pi>

The GPU provides Open GL ES 2.0, hardware-accelerated Open VG, and 1080p30 H.264 high-profile decode and is capable of 1Gpixel/s, 1.5Gtexel/s or 24 GFLOPs of general purpose compute. What's that all mean? It means that if you plug the Raspberry Pi 3 into your HDTV, you could watch BluRay quality video, using H.264 at 40Mbits/s



The biggest change that has been enacted with the Raspberry Pi 3 is an upgrade to a next generation main processor and improved connectivity with Bluetooth Low Energy (BLE) and BCM43143 Wi-Fi on board. Additionally, the Raspberry Pi 3 has improved power management, with an upgraded switched power source up to 2.5 Amps, to support more powerful external USB devices.



<http://uk.farnell.com/buy-raspberry-pi>

<http://www.newark.com/buy-raspberry-pi>

The Raspberry Pi 3's four built-in USB ports provide enough connectivity for a mouse, keyboard, or anything else that you feel the RPi needs, but if you want to add even more you can still use a USB hub. Keep in mind, it is recommended that you use a powered hub so as not to overtax the on-board voltage regulator. Powering the Raspberry Pi 3 is easy, just plug any USB power supply into the micro-USB port. There's no power button so the Pi will begin to boot as soon as power is applied, to turn it off simply remove power. The four built-in USB ports can even output up to 1.2A enabling you to connect more power hungry USB devices (This does require a 2Amp micro USB Power Supply)



On top of all that, the low-level peripherals on the Pi make it great for hardware hacking. The 0.1" spaced 40-pin GPIO header on the Pi gives you access to 27 GPIO, UART, I²C, SPI as well as 3.3 and 5V sources. Each pin on the GPIO header is identical to its predecessor the Model B+.

SoC

Built specifically for the new Pi 3, the Broadcom BCM2837 system-on-chip (SoC) includes four high-performance ARM Cortex-A53 processing cores running at 1.2GHz with 32kB Level 1 and 512kB Level 2 cache memory, a VideoCore IV graphics processor, and is linked to a 1GB LPDDR2 memory module on the rear of the board.



GPIO

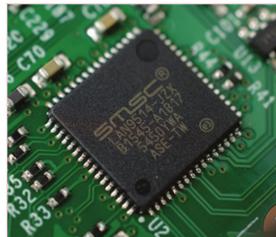
The Raspberry Pi 3 features the same 40-pin general-purpose input-output (GPIO) header as all the Pis going back to the Model B+ and Model A+. Any existing GPIO hardware will work without modification; the only change is a switch to which UART is exposed on the GPIO's pins, but that's handled internally by the operating system.



Pin#	NAME	NAME	Pin#
01	3.3v DC Power	DC Power 5v	02
03	GPIO2 (SDA1, I2C)	DC Power 5v	04
05	GPIO3 (SCL1, I2C)	Ground	06
07	GPIO4 (GPIO_GCLK)	(TXD0) GPIO14	08
09	Ground	(RXD0) GPIO15	10
11	GPIO17 (GPIO_GEN0)	(GPIO_GEN1) GPIO18	12
13	GPIO27 (GPIO_GEN2)	Ground	14
15	GPIO22 (GPIO_GEN3)	(GPIO_GEN4) GPIO23	16
17	3.3v DC Power	(GPIO_GEN5) GPIO24	18
19	GPIO10 (SPI_MOSI)	Ground	20
21	GPIO9 (SPI_MISO)	(GPIO_GEN6) GPIO25	22
23	GPIO11 (SPI_CLK)	(SPI_CEL_N) GPIO18	24
25	Ground	(SPI_CEL_N) GPIO17	26
27	ID_SD (I2C ID EEPROM)	(I2C ID EEPROM) ID_SC	28
29	GPIO15	Ground	30
31	GPIO16	GPIO12	32
33	GPIO13	Ground	34
35	GPIO19	GPIO16	36
37	GPIO26	GPIO20	38
39	Ground	GPIO21	40

USB chip

The Raspberry Pi 3 shares the same SMSC LAN9514 chip as its predecessor, the Raspberry Pi 2, adding 10/100 Ethernet connectivity and four USB channels to the board. As before, the SMSC chip connects to the SoC via a single USB channel, acting as a USB-to-Ethernet adaptor and USB hub.



Antenna

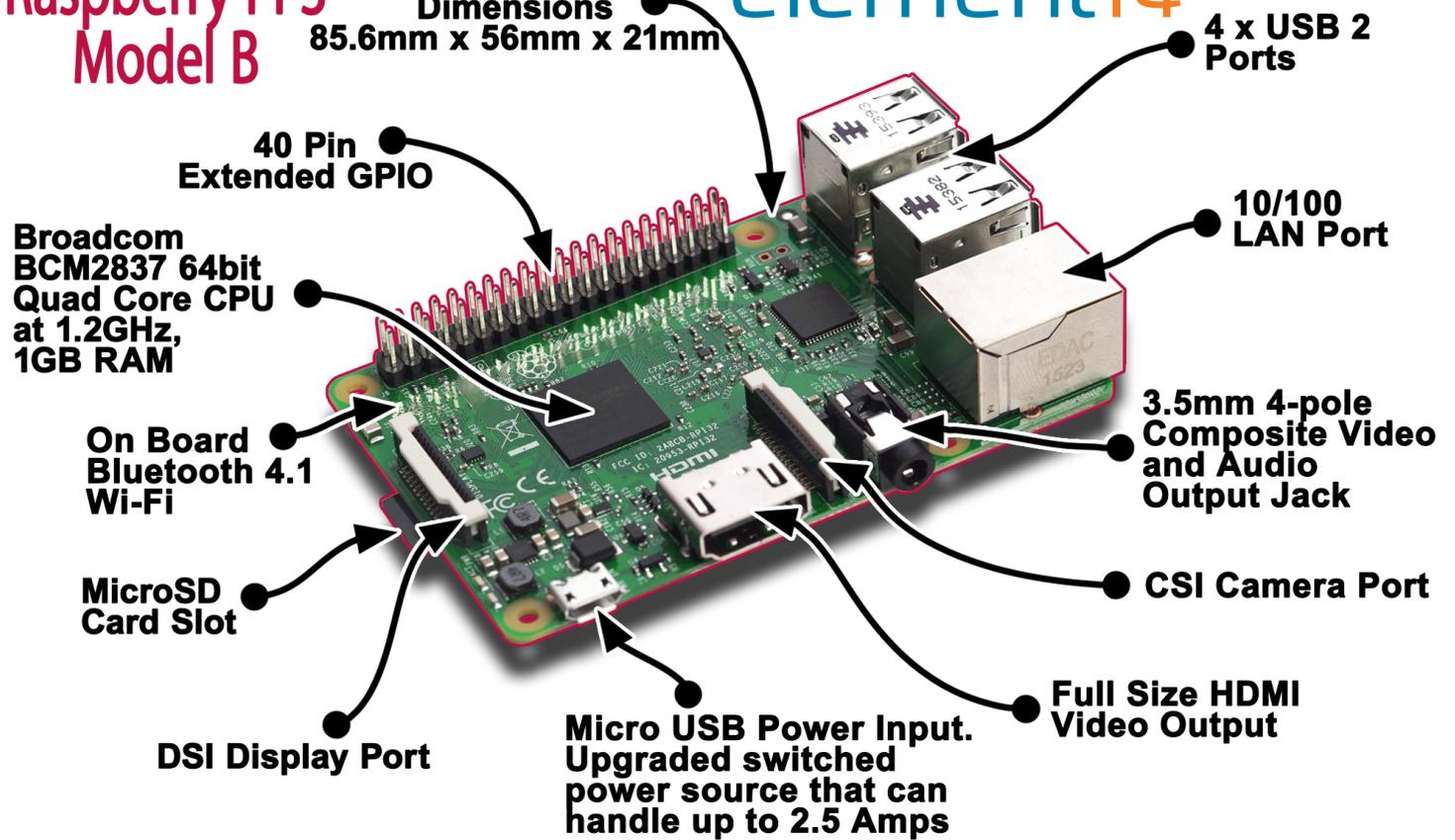
There's no need to connect an external antenna to the Raspberry Pi 3. Its radios are connected to this chip antenna soldered directly to the board, in order to keep the size of the device to a minimum. Despite its diminutive stature, this antenna should be more than capable of picking up wireless LAN and Bluetooth signals – even through walls.



Raspberry Pi 3 Model B

Dimensions
85.6mm x 56mm x 21mm

element14



Key Improvements from Pi 2 Model B to Pi 3 Model B:

- Next Generation QUAD Core Broadcom BCM2837 64bit ARMv7 processor
- Processor speed has increased from 900MHz on Pi 2 to 1.25Ghz on the RPi 3 Model B
- BCM43143 Wi-Fi on board
- Bluetooth Low Energy (BLE) on board
- Upgraded switched power source up to 2.5 Amps (can now power even more powerful devices over USB ports)

The main differences are the quad core 64-bit CPU and on-board Wi-Fi and Bluetooth. The RAM remains 1GB and there is no change to the USB or Ethernet ports. However, the upgraded power management should mean the Pi 3 can make use of more power hungry USB devices

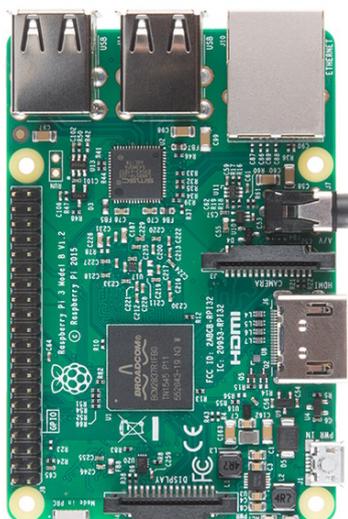
For Raspberry Pi 3, Broadcom have supported us with a new SoC, BCM2837. This retains the same basic architecture as its predecessors BCM2835 and BCM2836, so all those projects and tutorials which rely on the precise details of the Raspberry Pi hardware will continue to work. The 900MHz 32-bit quad-core ARM Cortex-A7 CPU complex has been replaced by a custom-hardened 1.2GHz 64-bit quad-core ARM Cortex-A53

In terms of size it is identical to the B+ and Pi 2. All the connectors and mounting holes are in the same place so all existing add-ons, HATs and cases should fit just fine although the power and activity LEDs have moved to make room for the WiFi antenna.

The performance of the Pi 3 is roughly 50-60% faster than the Pi 2 which means it is ten times faster than the original Pi.

All of the connectors are in the same place and have the same functionality, and the board can still be run from a 5V micro-USB power adapter. This time round, we're recommending a 2.5A adapter if you want to connect power-hungry USB devices to the Raspberry Pi.

Raspberry Pi 3 Model B



Raspberry Pi 2 Model B



<http://uk.farnell.com/buy-raspberry-pi>

<http://www.newark.com/buy-raspberry-pi>

	Raspberry Pi 3 Model B	Raspberry Pi 2 Model B	Model B+	Model A+	Model A	CMDK
Processor Chipset	Broadcom BCM2837 64Bit ARMv7 Quad Core Processor powered Single Board Computer running at 1250MHz	Broadcom BCM2836 32bit ARMv7 Quad Core Processor powered Single Board Computer running at 900MHz	Broadcom BCM2835 32bit ARMv6 SoC full HD multimedia applications processor	Broadcom BCM2835 32bit ARMv6 SoC full HD multimedia applications processor	Broadcom BCM2835 32bit ARMv6 SoC full HD multimedia applications processor	Broadcom BCM2835 32bit ARMv6 SoC full HD multimedia applications processor
GPU	Videocore IV	Videocore IV	Videocore IV	Videocore IV	Videocore IV	Videocore IV
Processor Speed	QUAD Core @1250 MHz	QUAD Core @900 MHz	Single Core @700 MHz	Single Core @700 MHz	Single Core @700 MHz	Single Core @700 MHz
RAM	1GB SDRAM @ 400 MHz	1GB SDRAM @ 400 MHz	512 MB SDRAM @ 400 MHz	256 MB SDRAM @ 400 MHz	256 MB SDRAM @ 400 MHz	512 MB SDRAM @ 400 MHz
Storage	MicroSD	MicroSD	MicroSD	MicroSD	SDCard	4GB eMMC
USB 2.0	4x USB Ports	4x USB Ports	4x USB Ports	1x USB Port	1x USB Port	1x USB Port
Power Draw / voltage	2.5A @ 5V	1.8A @ 5V	1.8A @ 5V	1.8A @ 5V	1.2A @ 5V	1.8A @ 5V
GPIO	40 pin	40 pin	40 pin	40 pin	26 pin	120 pin
Ethernet Port	Yes	Yes	Yes	No	No	No
Wi-Fi	Built in	No	No	No	No	No
Bluetooth LE	Built in	No	No	No	No	No