

# Visión estéreo para personas no videntes con cámaras de bajo costo

**Neira, Rodolfo Eduardo**

[rodolfoneira8@gmail.com](mailto:rodolfoneira8@gmail.com)

**Lurgo, Gerardo Jorge**

[slurgo@arnet.com.ar](mailto:slurgo@arnet.com.ar)

**Rubiolo, Bruno**

[brunorubiolo1@gmail.com](mailto:brunorubiolo1@gmail.com)

**Burgos, Fabián Agustín**

[fabianburgos.09@gmail.com](mailto:fabianburgos.09@gmail.com)

*Universidad Tecnológica Nacional - Facultad Regional San Francisco (Argentina).*

Fecha de recepción: 05/07/2020

Fecha de aprobación COINI: 08/10/2020

Fecha de aprobación RIII: 22/06/2021

## **RESUMEN**

En este trabajo se presenta el avance del diseño de un equipo de visión estéreo con cámaras de bajo costo capaz de detectar objetos dentro de una habitación, para facilitar la movilidad de personas no videntes dentro de la misma, evitando que sufra accidentes con algún objeto presente. Con este equipo se trata de lograr una mejor integración social de la persona ciega, aumentando sus expectativas en cuanto a movilidad y libertad de acción. Se ha demostrado que estas tecnologías favorecen la participación de la persona ciega en el medio familiar cercano y, también que posee un importante alcance psicológico desde lo emocional. Además, aumenta las posibilidades que tienen las instituciones relacionadas con esta temática al contar con tecnologías activas que posibilitan un mayor alcance para la atención efectiva de las personas que asisten. El uso de este equipo disminuye, de manera significativa los accidentes que se puedan ocasionar cuando al trasladarse la persona no vidente choque con objetos existentes y pueda producirse lesiones graves. De este modo la problemática de la persona con discapacidad visual se la aborda de un modo más inclusivo, donde no solo se centra en ella la atención de manera exclusiva, sino que también incluye al grupo familiar más cercano y, a las instituciones que tratan esta situación.

**Palabras Claves:** Discapacidad Visual, Integración Social, Persona, Bajo costo, Cámaras

## **Stereo vision for blind people with low cost cameras**

### **ABSTRACT**

This work presents the advancement of the design of a stereo vision equipment with low-cost cameras capable of detecting objects within a room, to facilitate the mobility of blind people within it, avoiding accidents with an object present. With this equipment, the aim is to achieve a better social integration of the blind person, increasing their expectations in terms of mobility and freedom of action. It has been shown that these technologies favor the participation of the blind person in the close family environment and also that it has an important psychological scope from the emotional point of view. In addition, it increases the possibilities that institutions related to this subject have by having active technologies that allow a greater scope for the effective care of the people who attend. The use of this equipment significantly reduces accidents that may occur when the blind person collides with existing objects when moving and serious injuries may occur. In this way, the problem of the person with visual impairment, is approached in a more inclusive way, where not only attention is focused on them exclusively, but also includes the closest family group and the institutions that treat this situation.

**Keywords:** Visual disability, Social integration, Person, Low cost, Cameras

## **Visão estérea para pessoas que não enxergam com câmeras baixo custo**

### **RESUMO**

Neste trabalho se apresenta o avanço do desenho de um aparelho de visão estérea com câmeras de baixo custo capaz de detectar objetos dentro de uma habitação, para facilitar a mobilização de pessoas não videntes dentro da mesma, evitando que sofra acidentes com algum objeto presente. Com este aparelho se trata de conseguir uma melhor integração social da pessoa cega, aumentando suas expectativas em quanto a mobilização e liberdade de ação. Tem se demonstrado que estas tecnologías favorecem a participação da pessoa cega no meio familiar mais próximo e, também que tem um importante alcance psicológico desde o emocional. Além disso, aumenta as possibilidades que têm as instituições relacionadas com esta temática ao contar com tecnologías ativas que possibilitam um maior alcance para a atenção efetiva das pessoas que assistem. O uso deste aparelho diminui, de maneira significativa os acidentes que se possam ocasionar quando ao se trasladar a pessoa que não enxerga bata com objetos existentes e possa se produzir lesões graves. Deste modo a problemática da pessoa com deficiência visual se aborda de um modo mais inclusivo, onde não só se centra nela a atenção de maneira exclusiva, se não que também incluye ao grupo familiar mais próximo e, às instituições que tratam esta situação.

**Palavras chave:** Deficiência visual, Integração social, Pessoa, Baixo custo, Câmeras

## 1. INTRODUCCIÓN

El desarrollo de cámaras digitales con alta tecnología y día a día con mejores capacidades, ha permitido que en diferentes aplicaciones se utilicen como una alternativa a la visión humana, en diferentes tareas como, en la supervisión de personas mediante sistemas inteligentes que pueden detectar, contar, identificar y seguir la trayectoria de las personas [1], [2]; en sistemas de inspección visual automatizados de productos para el control de su calidad en las empresas de manufactura [3], entre muchas otras aplicaciones que crecen día a día.

El empleo de cámaras como sensores de medición de distancia ha tenido diferentes aplicaciones. Por ejemplo, su utilización para permitir la navegación autónoma de robots terrestres [4]; en arquitectura como un instrumento de medición en interiores para obtener las dimensiones de paredes y pisos, así como para la ubicación correcta de muebles con el propósito de diseño de interiores, en exteriores para medir el tamaño y la posición de ventanas y puertas [5], y muchas más. Una imagen o secuencia de imágenes trae consigo una cantidad muy grande de información geométrica acerca de la escena representada, se han desarrollado diferentes técnicas para la construcción de escenarios 3D a partir de imágenes en 2D [6].

Se han desarrollado técnicas y métodos de descomposición de imágenes para su representación en el espacio del mundo real. Thomas Bucher [7], describe un método para mapear una imagen a coordenadas del mundo real y obtener así, una aproximación de la altura de objetos, longitudes y cambios de posición; basándose en un pequeño grupo de parámetros de fácil estimación a partir de características de los objetos o marcas en la escena, esto sin la necesidad de requerir alguno de los parámetros intrínsecos de la cámara.

Lázaro et al. [8], presentan la caracterización de la variación de intensidad de niveles de grises y su análisis mediante FFT (Fast Fourier Transform: transformada rápida de Fourier), en imágenes tomadas para medir la distancia entre un diodo emisor de infrarrojo y el centro de una cámara. El método propuesto se aplicó para hacer una estimación de distancias en el rango de 420 a 800 cm, logrando una exactitud sobre el 3%.

Un campo en el que confluyen cámaras y ordenadores es, como su nombre indica, el de la Visión por Computador y dentro de éste la visión estereoscópica. Como fácilmente se puede deducir, las imágenes son bidimensionales mientras que la escena cotidiana es tridimensional. Esto significa que entre el paso de la escena, que es la realidad, a la imagen se ha perdido lo que denominamos la tercera dimensión. La visión estereoscópica constituye un procedimiento más para la obtención de esa tercera dimensión perdida y a partir de ella en la medida de lo posible la obtención de la forma de los objetos en la escena. Nuestro sistema visual humano es capaz de percibir en tres dimensiones y además es estereoscópico, constituido por dos ojos, ello es lo que ha hecho que los sistemas estereoscópicos artificiales utilicen al menos, dos imágenes distintas de la misma escena. Con ellas se puede llegar a determinar la distancia a la que se encuentra un objeto cualquiera, contenido en las dos imágenes, respecto del observador. Las cámaras se utilizan para captar las imágenes y el computador se requiere para realizar los cálculos que determinan la distancia al observador. En trabajos previos([9], [10]), se han desarrollado técnicas computacionales basadas en el análisis de imágenes estereoscópicas con el fin de determinar la estructura tridimensional de la escena captada y poder obtener parámetros tales como distancia entre los objetos de la imagen.

La visión estereoscópica está enfocada a analizar dos imágenes de la misma escena tomadas al mismo tiempo, con el objetivo de obtener diferentes puntos de vista de los objetos en la escena y en consecuencia más información de los objetos de análisis. Específicamente lo que comúnmente se busca con esta metodología es, el conocer la profundidad en la fotografía. Otro de los aspectos importantes que

nos aporta es la capacidad de medir las dimensiones de los objetos presentes en la escena, siempre y cuando dichos objetos se encuentren dentro de ambas imágenes del sistema y las cámaras se encuentren calibradas. Básicamente el problema en visión estereoscópica es: estimar un punto en el espacio de 3 dimensiones a partir de 2 fotografías distintas del punto deseado, todo con respecto a un sistema de referencia. Esto normalmente se resuelve por aproximación geométrica, comúnmente llamada triangulación.

## 2. MÉTODOS Y MATERIALES

En la actualidad es casi indispensable el uso de la tecnología para la solución a problemas cotidianos en la sociedad. Las cámaras digitales se utilizan en diversos ambientes y para distintas finalidades, destacando por ejemplo: seguridad (vigilancia de edificios, museos), comunicaciones (videoconferencias), entretenimiento (videoconsolas, cámaras fotográficas, cámaras de video), etc. Con la ayuda de una computadora se pueden desarrollar increíbles aparatos tecnológicos.

En este proyecto diseñaremos un equipo capaz de detectar objetos dentro de una habitación para incluir y facilitar la movilidad de personas no videntes dentro de la sala, de esta forma se les indica cuando están cerca de un objeto para evitar que la persona se accidente mediante un choque con el mismo. También se le indicaran las ubicaciones de las puertas de la habitación.

En la Figura 1 se observa el siguiente diagrama en bloques que resume el proyecto.

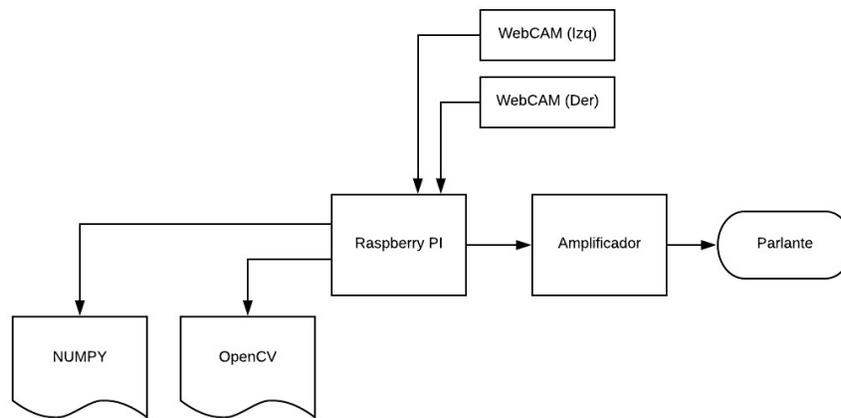


Figura 1 Diagrama General del proyecto.

### 2.1. Elección de la cámara

A partir de la problemática planteada anteriormente se parte desarrollando el equipo, en especial la parte de visión estéreo con cámaras de bajo costo. Luego de un estudio sobre las cámaras digitales que se comercializan en el mercado local, decidimos utilizar una cámara Web marca Logitech modelo C270 por su relación precio-calidad ya que se utilizarían dos, las mismas poseen una resolución HD 720p que nos permite trabajar bien sobre la imagen y, además nos brinda un refresco de imagen bastante adecuado para su utilización gracias a los 30 FPS de la misma, según lo indica la Figura 2.



Figura 2 Cámara Web.

Las características técnicas de la cámara Web marca Logitech modelo C270, se muestran en la Tabla 1

Tabla 1 Características técnicas.

Parámetro	Característica
Marca	Logitech
Modelo	C270
Línea	HD
Modelo alfanumérico	V-U0018
Micrófono Incorporado	Sí (mono)
Resolución de vídeo	HD 720p
Resolución de imagen	3 Mpx
Interfaces	USB 2.0
Sistema operativo que soporta	Windows 7, Windows Vista, Windows XP, Mac, Windows 10, Windows 8.1, Windows 8
Fluidez de video	30 FPS
Protección de imagen	Sí
Tipo de enfoque	Foco fijo
Tecnología de lente	Estándar
Campo visual	60°

## 2.2. Distorsión en las imágenes

Las cámaras de bajo costo existentes en el mercado actual introducen mucha distorsión en las imágenes. Dos distorsiones principales son la distorsión radial y la distorsión tangencial.

Debido a la distorsión radial, las líneas rectas aparecerán curvadas. Su efecto es mayor a medida que nos alejamos del centro de la imagen, como se puede observar en la Figura 3. Se puede ver que el borde no es una línea recta y no coincide.

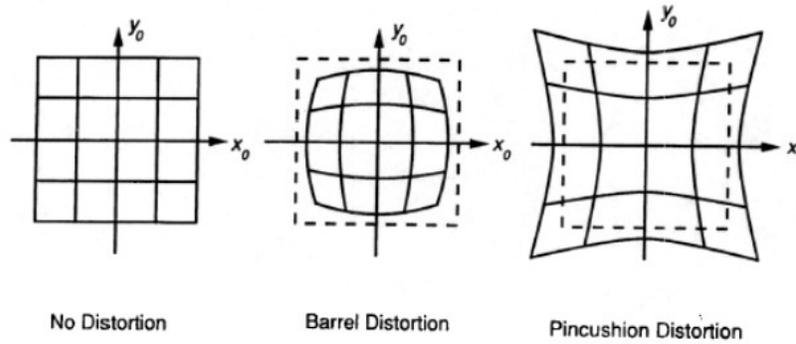


Figura 3 Distorsión de las imágenes.

Esta distorsión se resuelve de la siguiente manera, mediante la Ecuación (1) y la Ecuación (2):

$$x_{corrected} = x(1 + k_1r^2 + k_2r^4 + k_3r^6) \quad (1)$$

$$y_{corrected} = y(1 + k_1r^2 + k_2r^4 + k_3r^6) \quad (2)$$

Del mismo modo, otra distorsión es la tangencial que ocurre porque la toma de imágenes no está perfectamente alineada paralelamente al plano de imagen. Por lo tanto, algunas áreas en la imagen pueden verse más cercanas de lo esperado. Se resuelve como sigue, con la Ecuación (3) y la Ecuación (4):

$$x_{corrected} = x + [2p_1xy + p_2(r^2 + 2x^2)] \quad (3)$$

$$y_{corrected} = y + [p_1(r^2 + 2y^2) + 2p_2xy] \quad (4)$$

En resumen, necesitamos encontrar cinco parámetros, conocidos como coeficientes de distorsión dados, como se observa en la Ecuación (5):

$$Distortion\ coefficients = (k_1 \ k_2 \ p_1 \ p_2 \ k_3) \quad (5)$$

Además de los coeficientes de distorsión dados, necesitamos encontrar más información, como los parámetros intrínsecos y extrínsecos de una cámara digital. Los parámetros intrínsecos son específicos de una cámara. Incluye información como distancia focal ( $f_x, f_y$ ), centros ópticos ( $c_x, c_y$ ), etc. También se llama matriz de cámara. Depende sólo de la cámara, por lo que una vez calculada, se puede almacenar para fines futuros. Se expresa como una matriz de 3x3 (6):

$$\text{camera matrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

Los parámetros extrínsecos corresponden a vectores de rotación y traslación que traducen las coordenadas de un punto 3D a un sistema de coordenadas.

Para las aplicaciones estéreo, estas distorsiones necesitan ser corregidas primero. Para encontrar todos estos parámetros, lo que tenemos que hacer es proporcionar algunas imágenes de muestra de un patrón bien definido (tablero de ajedrez), como se observa en la Figura 4. Encontramos algunos puntos específicos en él (esquinas cuadradas en tablero de ajedrez). Conocemos sus coordenadas en el espacio del mundo real y conocemos sus coordenadas en imagen. Con estos datos, un programa matemático lo resuelve para obtener los coeficientes de distorsión. Para obtener mejores resultados, necesitamos al menos 10 patrones de prueba.

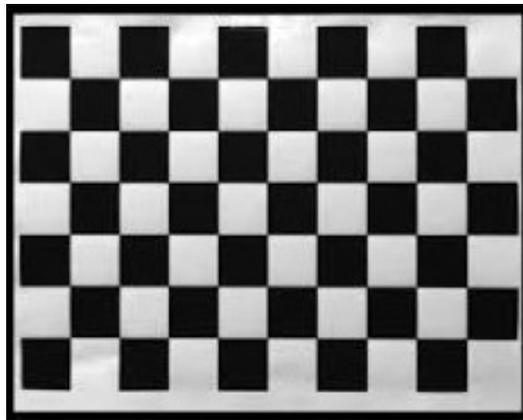


Figura 4 Vista de tablero de ajedrez.

### 2.3. Desarrollo del código de programación

Para realizar lo mencionado en este proyecto se desarrolló una programación en Python 2.7 mediante el editor de texto NOTEPAD++ unido con la librería OPENCV y NUMPY. La primera librería facilita el procesamiento de la imagen como lo es el escalado de colores y la generación de los mapas de disparidad, mientras que la última librería mencionada nos facilita el cálculo numérico y de matrices.

OpenCV es una librería de código abierto altamente optimizada para aplicaciones de visión artificial (en tiempo real), desarrollada por la firma Intel en 1999, y en la actualidad es utilizada como una herramienta en aplicaciones tales como: reconstrucción 3D a partir de imágenes, reconocimiento de objetos, detección de movimiento, entre otras.

Dentro de sus principales características, se mencionan su libre disponibilidad, como así también su posibilidad de usarse en varias plataformas, tales como: Windows, Mac OS X, GNU/Linux, Android, para arquitecturas de hardware ARM (dispositivos móviles y Raspberry PI) y para PC.

Es muy versátil para aprovechar la potencia de procesamiento de las placas gráficas, lo que permite el tratamiento de imágenes binarias, en tono de grises y a color, según Figura 5.

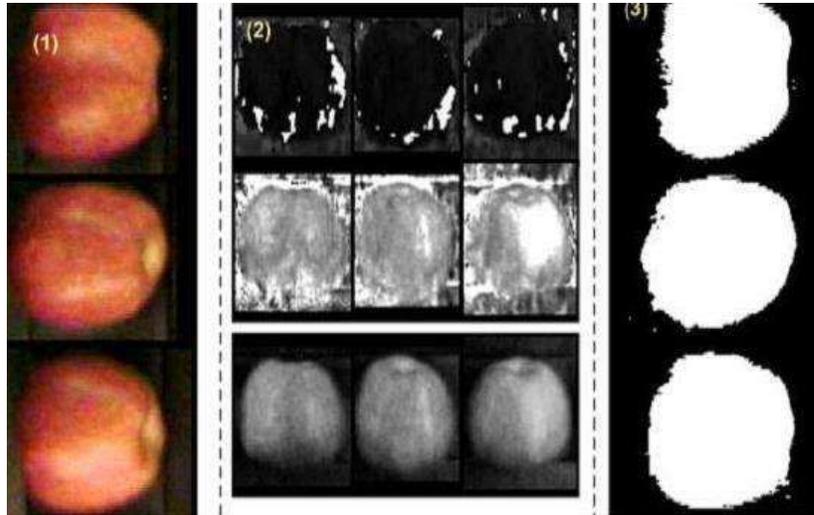


Figura 5 Vista de imagen usando OpenCV.

Al utilizar el lenguaje de programación *Python* para analizar datos, considerando que al final casi todo se reduce a realizar cálculos numéricos con matrices de dimensiones considerables, NumPy se convierte entonces en una herramienta esencial.

NumPy es un paquete de Python que significa “Numerical Python”, es la librería principal para la informática científica, proporciona potentes estructuras de datos, implementando matrices y matrices multidimensionales. Estas estructuras de datos garantizan cálculos eficientes con matrices. Los beneficios que aporta son: ser más compacto, acceder más rápido a leer y escribir artículos, ser más conveniente y más eficiente.

Algunos detalles técnicos, que fueron tenidos en cuenta para su aplicación, se mencionan a continuación:

1. Es distinta, y mucho más eficiente, la manera en que se accede a los elementos de un array de NumPy con respecto a cómo Python procede a realizar tal tarea en sus estructuras de datos básicas,
2. El número de comprobaciones intermedias a la hora de llevar a cabo cálculos numéricos es menor en NumPy, y
3. NumPy está escrito utilizando el lenguaje de programación C, que es bastante más rápido que Python.

Para el desarrollo del equipo, la intención del proyecto es utilizar el ejecutador de Python 2.7 en una computadora u ordenador tal como una Raspberry Pi 3 model B+, de modo que el sistema sea compacto, eficiente, de bajo consumo, confiable y económico. Esta placa realizará el procesamiento de las cámaras mediante el código en lenguaje Python y advertirá a la persona no vidente sobre los obstáculos y puertas presentes en los ambientes cerrados.

En la etapa inicial del proyecto se desarrolló el código de programación en un ordenador portátil con Windows 10, debido a su versatilidad de modificación, dicho código es capaz de detectar objetos en movimientos.

El sistema deberá avisar a la persona no vidente cuando la distancia del objeto sea menor o igual a la establecida y, para eso se debe medir la distancia entre las cámaras para luego encontrar la expresión que nos facilitará calcular la distancia a un objeto. Esta también depende de la distancia focal propia de la cámara digital, pixeles, y ángulo de apertura, tal como se indica en la Figura 6.

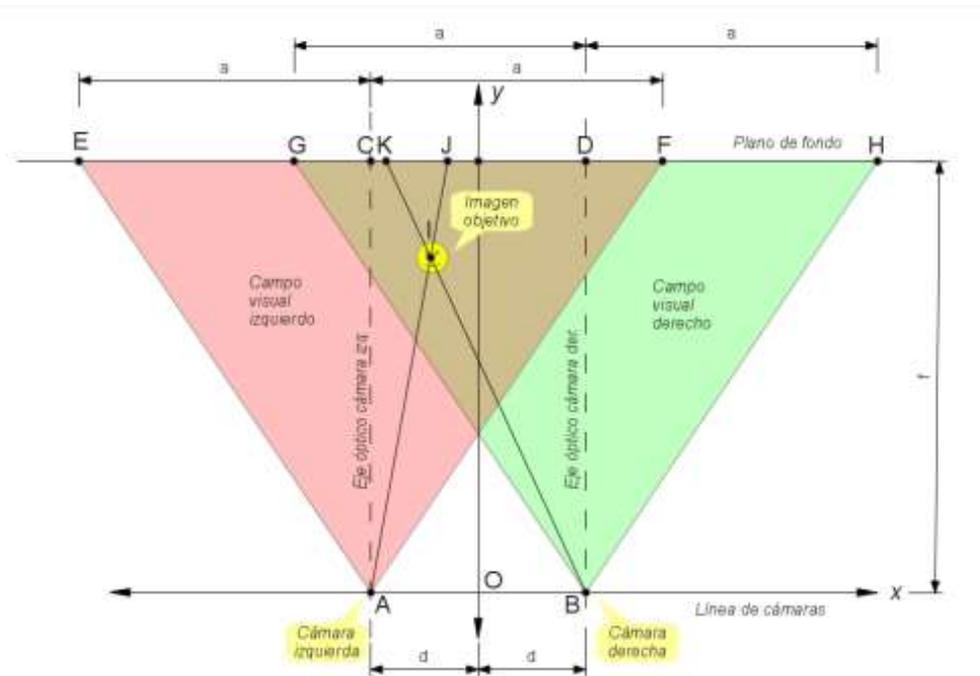


Figura 6 Vista de la distancia entre las cámaras digitales.

En la imagen superior se puede apreciar las distancias de las cámaras y la triangulación necesaria para obtener la distancia entre el objeto y el centro (ubicado en el medio de las dos cámaras).

A partir de dos imágenes (una de la cámara izquierda y una de la cámara derecha) se procede a detectar el mismo objeto, y mediante la fórmula matemática detectar la distancia al objeto, tal como se indica en el ejemplo de la Figura 7.

#### 2.4. Desarrollo del proyecto

Con una fórmula matemática se compara dicha distancia con la distancia patrón y de esta forma se alertará a la persona no vidente cuando el obstáculo esté cerca para que pueda tomar las decisiones adecuadas.

Para hacer el proyecto viable como se mencionó anteriormente, se utilizará una Raspberry Pi 3 model B+ como procesamiento según Figura 8; para esto es necesario instalar una imagen de Linux y, luego instalar Python con el siguiente comando desde el símbolo del sistema: **sudo apt-get install python**

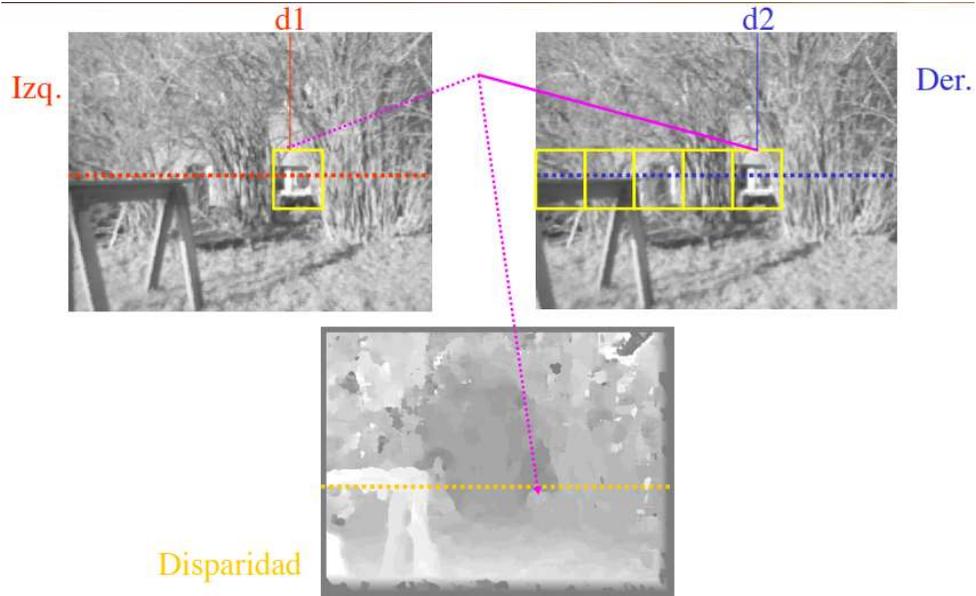


Figura 7 Imágenes de las cámaras digitales y su disparidad.

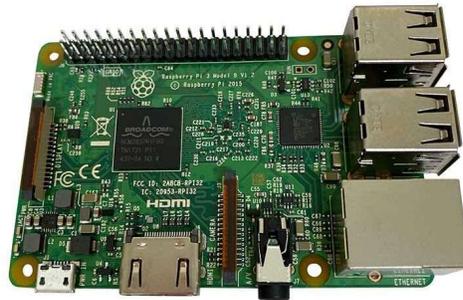


Figura 8 Vista de placa.

Las características técnicas de la placa Raspberry Pi 3 modelo B+, se muestran en la Tabla 2

Tabla 2 Características técnicas.

<b>Parámetro</b>	<b>Característica</b>
CPU	ARMv8 1.2 GHz 64-bit quad-core
Modelo	802.11n Wireless LAN
Protocolo de comunicación	Bluetooth 4.1
Tecnología de red	Bluetooth Low Energy (BLE)
Memoria RAM	1 GB
Puertos USB	4
pinos GPIO	40
Puerto	Puerto full HDMI

Sistema operativo que soporta	Windows 10, Linux, incluyendo Snappy Ubuntu Core
Puerto de red	Ethernet
Audio	Audio jack de 3.5 mm combinado con salida de video compuesto
Interfaz de display	DSI
Tarjeta	Slot Micro SD
Tarjeta de video	Video Core IV 3D graphics core

Para lograr el inicio automático del script de Python se utiliza el crontab de Linux de modo tal que al alimentar el circuito, se ejecutará de forma autónoma el script que contiene la programación de detección de obstáculos. Para configurar esto se utilizan las siguientes líneas de comandos desde el símbolo del sistema:

```
SUDO crontab -e

@reboot Python /home/pi/Desktop/Script.py

Control+S

Control+x

sudo /etc/init.d/rsyslog restart
```

Dicho script utiliza los pines GPIO de la placa Raspberry para emitir el sonido de salida, estos pines manejan corrientes bajas del orden de los 2mA y voltajes en formato ttl de 3.3V, según Figura 9.

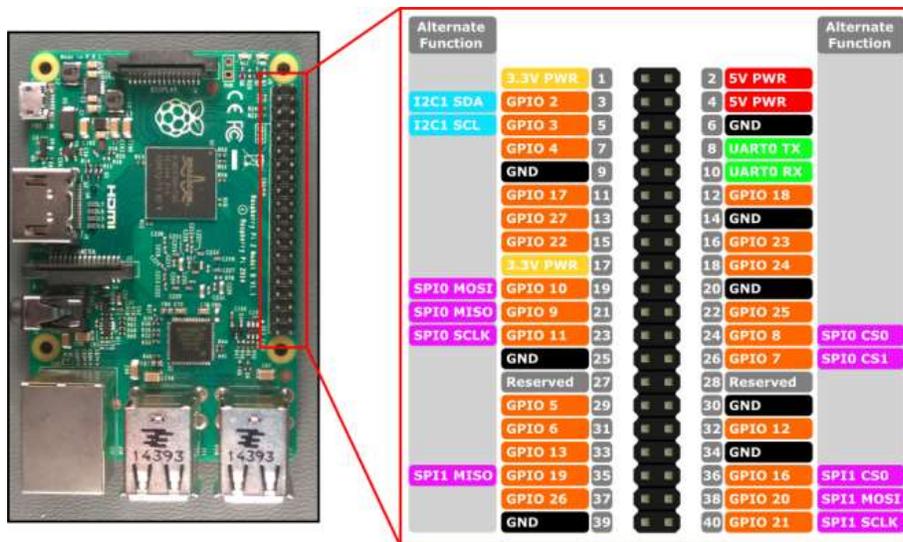


Figura 9 Vista de placa y conectores GPIO.

Para lograr emitir el sonido en un parlante se utiliza un amplificador configurado de la siguiente manera, como se indica en la Figura 10.

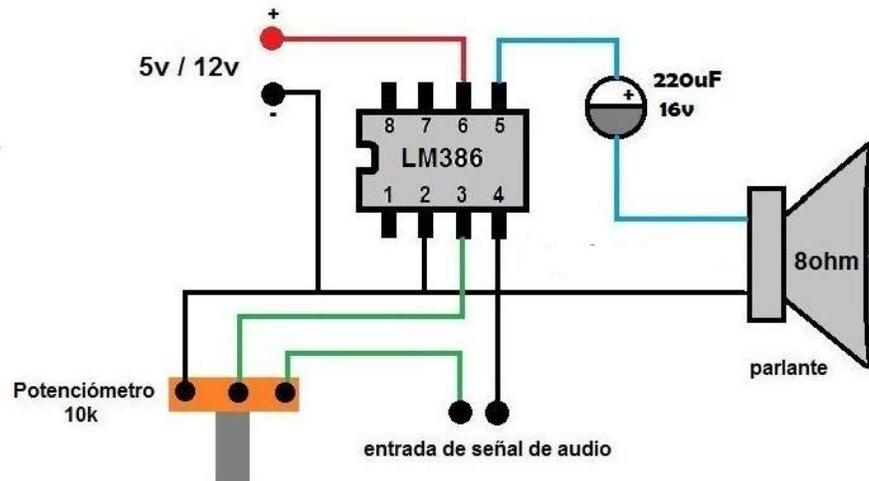


Figura 10 Vista del amplificador del parlante.

El LM386 es un circuito integrado que consiste en un amplificador que requiere bajo voltaje, tanto en la entrada de audio como en la alimentación. Suministrando 12 Volt se puede obtener 0,5 W de potencia, con solo un 0,2% de distorsión.

#### 4. RESULTADOS

Mediante un estudio de los elementos existentes aplicados a facilitar la movilidad de personas no videntes, evitando que sufran accidentes con algún objeto presente, dentro de ambientes cerrados como una casa de familia o una oficina, hemos observado que los mismos son escasos y, de un costo elevado, generalmente por su procedencia del exterior.

Con el diseño y construcción de este equipo se logra mejorar la integración de la persona no vidente, así como disminuir notablemente los inconvenientes que origina en los familiares directos, reduciendo tiempos improductivos de quien lo asiste o acompaña y, mejorando su calidad de vida al tener mayor autonomía para desenvolverse en su hogar o en otro ámbito donde desarrolle sus actividades.

La construcción de este equipo, es producto de un desarrollo local que traería muchos beneficios al aprovechar la capacidad técnica e intelectual de la comunidad y, permitiría su aplicación a nivel regional y nacional.

#### 5. CONCLUSIONES

El sentido de la vista es uno de los más importantes del ser humano, porque a través de ellos percibimos la mayor parte de la información del mundo que nos rodea.

Con los avances en tecnología de cámaras digitales que presentan mejores capacidades de resolución, ha permitido que se utilicen como una alternativa válida a la visión humana, lo que posibilita a las personas con discapacidad visual, desarrollar muchas tareas productivas y, ser un ejemplo para los demás, logrando su integración a la sociedad pero principalmente, que se sientan capaces de hacer lo que se propongan.

El acceso a estas nuevas tecnologías puede resultar algo costoso para algunas personas, pero cada vez que ingresa al mercado un producto nuevo, los costos van disminuyendo y, se vuelven accesibles cada vez más para la sociedad toda.

Con el desarrollo de este equipo, se obtendrá información que puede ser utilizada para mejorar en forma significativa la construcción de ambientes adaptados con tecnología disponible para personas no videntes, además aportar información importante para potenciar políticas públicas para mejorar la movilidad de personas con baja visión o ceguera, entre otros avances.

Además, articular experiencias con organismos que trabajan la problemática de la discapacidad visual, se retroalimentarán mutuamente pudiendo utilizar herramientas estadísticas para ser aplicadas en la mejora de la calidad de vida de las mismas.

## REFERENCIAS.

- [1] Dee, H. M., Velastin, S. A. (2008) How close are we to solving the problem of automated visual surveillance? *Machine Vision and Applications*, vol. 19, no. 5-6, pp. 329-343.
- [2] Vera P., Zenteno, D., Salas, J. (2013) Counting Pedestrians in Bidirectional Scenarios Using Zenithal Depth Images, *Proceedings of 5th Mexican Conference, MCPR, Querétaro, Mexico*. Junio 26-29, pp. 84-93.
- [3] Chin, R. T., Harlow, C. A. (1982) Automated visual inspection: A survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 557-573.
- [4] Royer E. et al., (2007) Monocular vision for mobile robot localization and autonomous navigation, *International Journal of Computer Vision*, vol. 74, no. 3, pp. 237-260.
- [5] Criminisi, A., Reid, I., Zisserman, A. (1999) A plane measuring device, *Image and Vision Computing*, vol. 17, no. 8, pp. 625-634.
- [6] Wan Y. et al., (2012) A Study in 3D-Reconstruction Using Kinect Sensor, *8th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, IEEE, pp. 1-7.
- [7] Bucher, T. (2000) Measurement of distance and height in images based on easy attainable calibration parameters, *Proceedings of the IEEE in Intelligent Vehicles Symposium, IV*, pp. 314-319.
- [8] Lázaro J. L. et al. (2009) Sensor for distance estimation using FFT of images. *Sensors*, vol. 9, no. 12, pp. 10434-10446.
- [9] Alvarez L., Sanchez J., Sobre algunos problemas reales en visión por ordenador que conducen a sistemas de ecuaciones algebraicos., *Departamento de Informática y Sistemas, Universidad de Las Palmas de G.C.*
- [10] Brady M., Wang H., (1992). *Visión for Mobile Robots*, Robotics Research Group, Department of Engineering Science, University of Oxford.