



UNIÓN DE ASOCIACIONES
DE INGENIEROS TÉCNICOS
INDUSTRIALES Y GRADUADOS
EN INGENIERÍA DE LA
RAMA INDUSTRIAL DE ESPAÑA

UNIÓN DE ASOCIACIONES DE INGENIEROS TÉCNICOS INDUSTRIALES Y GRADUADOS EN INGENIERÍA DE LA RAMA INDUSTRIAL DE ESPAÑA (UAITIE)

“CONVOCATORIA 2022”

VI PREMIO NACIONAL DE INICIACIÓN A LA INVESTIGACIÓN TECNOLÓGICA

Newsight

AUTORES:

Héctor Tablero, Héctor Polo, Alonso González, Pablo Romeo y Rodrigo
González

BLOQUE TEMÁTICO:

Atención a la discapacidad

NIVEL EDUCATIVO:

2º de Bachillerato

COORDINADOR:

Javier López Iglesias

Abril de 2022



UAITIE
Premio nacional de iniciación
a la investigación tecnológica
Convocatoria 2022

NEWSIGHT

Atención a la discapacidad



H U M Δ N I T A S

BILINGUAL SCHOOL TRES CANTOS

Héctor Polo Alonso González Pablo Romeo
Rodrigo González Héctor Tablero

2º de Bachillerato

RESUMEN

Nuestro proyecto consiste en la creación de unas gafas a las que hemos decidido llamar Newsight. El nombre hace referencia a su propósito: Newsight será una herramienta de orientación y adaptación para personas invidentes. Las gafas contarán con diversas funciones: capacidad para reconocer obstáculos y describir el entorno, lectura de texto a través del audífono incorporado en la patilla de la gafa, un sistema con reconocimiento de voz que permitirá al usuario convertir su habla en los comandos que dirigirán el funcionamiento del producto. Newsight también contará con un buscador de respuestas (Wolfram Alpha) que le dará al cliente la posibilidad de realizar cualquier tipo de preguntas a las gafas, y estas se encargarán de buscar y dictar las respuestas correspondientes.

La base del sistema operativo de las gafas está programada en Python, y adicionalmente hace uso de librerías escritas en C++. El programa de reconocimiento de voz funciona con la API de Google, el habla con el motor pyttsx3, la búsqueda de información a través de Wolfram Alpha, la lectura de texto valiéndose del módulo cv2 de OpenCV y PyTesseract, la clasificación se llevará a cabo empleando librerías de Microsoft Azure y finalmente la traducción se valdrá de las APIs de Google Translate y DeepL.

Los materiales que compondrán las gafas son bastante simples: un audífono, una cámara, un micrófono, estructura básica de unas gafas y dos baterías. Nuestro diseño aboga por la comodidad y el minimalismo.

Cabe recalcar que Newsight está orientado hacia las posibilidades, es decir; Newsight pretende ser un boceto de lo que se puede lograr con mayor tiempo y financiación.

ÍNDICE

RESUMEN	2
ÍNDICE	3
INTRODUCCIÓN	4
PALABRAS CLAVE	4
DESARROLLO	5
Etapa de pensamiento	5
Etapa de programación	5
Reconocimiento de voz y habla	5
Búsqueda de información	6
Lectura de texto	6
Calificación de imágenes	7
Traducción	7
Etapa de maquetación	7
OBJETIVOS DE NEWSIGHT	9
CONCLUSIONES	10
BIBLIOGRAFÍA	11

INTRODUCCIÓN

Nuestra percepción del ambiente determina cómo interactuamos con lo que nos rodea. Para las personas invidentes el día a día se convierte en un desafío a enfrentar. Newsight se encargaría de adaptar el ambiente a ellos a través de tecnología de reconocimiento de imagen. No está planteado como el sustituto al bastón tradicional o al perro guía, sino para ser una herramienta más a disposición de las personas con discapacidad visual. Sin embargo, Newsight no se quedaría en una única tarea. También cuenta con funciones como: lectura de texto y dictado de este mismo a través del audífono incorporado en las patillas de las gafas, indicación de fecha y hora, reconocimiento del tiempo atmosférico y función consultiva a través del buscador de respuestas Wolfram Alpha, que permitirá al usuario acceder a cualquier tipo de información.

Newsight funciona con un software programado en Python, el cual se encarga de reconocer las peticiones del cliente y transformarlas en comandos para que las gafas puedan realizar las acciones solicitadas.

Nuestro grupo considera nuestros sentidos un tesoro y nos resultaba injusto que algunos hayan sido privados de estos. Por tanto, nos planteamos la pregunta: ¿En qué podemos ayudar? Así surgió la idea de capacitar a aquellos que no pueden ver con un ojos artificiales.

PALABRAS CLAVE

Red neuronal¹, inteligencia artificial², API³, invidentes⁴

¹ Conjunto de unidades denominadas neuronas, que se transmiten señales entre sí y evolucionan para perfeccionar la red

² Aplicaciones informáticas que tratan de replicar el pensamiento humano

³ Conjunto de funciones usadas para integrar el software de una librería

⁴ Personas carentes del sentido de la vista

ANÁLISIS DEL PROBLEMA

Alrededor de 1 millón de personas en España sufren de ceguera y, de estos, unos 60.000 se encuentran en un estado de incapacidad visual total. Consideramos que esta minoría se encuentra marginada de los avances tecnológicos modernos. Para empezar, los métodos tradicionales como el bastón blanco o el perro guía se llevan empleando desde los 2000 cuando la Fundación Once del Perro Guía (FOPG) comenzó a distribuirlos. Desde entonces los avances que ha tenido este sector son casi inexistentes. Hoy en día contamos con múltiples herramientas de orientación (Google Maps, GPS, etc.) que están a disposición de todo el mundo excepto los miembros de este colectivo.

Por consiguiente hemos optado por implementar los avances nosotros mismos con la creación de un producto que se ajuste a las necesidades básicas de las personas que sufren de invidencia. Sin significar un óbice para su comodidad, su poder adquisitivo y su adaptación a la actualidad.

DESARROLLO

Etapa de pensamiento

Nuestro equipo tomó decisiones en base a las temáticas ofrecidas por nuestro profesorado, y consideró las que tendrían mayor impacto en la sociedad. En vez de centrarnos en simples mejoras destinadas a una población general, decidimos hacer una revolución en la calidad de vida de un colectivo que no recibe tanta atención como deberían. Poniendo enfoque en la revolución tecnológica que ha traído consigo la creación de redes neuronales artificiales y aprendizaje automático en máquinas, decidimos aplicar dichas innovaciones en nuestro producto, no solo para crear un dispositivo que aumente la comodidad de personas invidentes, sino también para demostrar las posibilidades que tiene este concepto, y lo que se podría llegar a hacer con más tiempo y recursos.

Tras analizar las posibilidades que podrían alcanzar diversos proyectos decidimos orientarnos hacia el entorno de las minusvalías. Como mencionamos anteriormente, nuestro grupo considera los sentidos el pináculo de los procesos evolutivos. A la postre determinamos centrarnos en la ceguera. De ahí, tras una breve consulta grupal, nos surgió la idea de Newsight; las gafas capaces de situar al invidente en un ambiente idóneo.

Etapa de programación

Ya que las acciones que Newsight hace de forma automática están limitadas, el programa está diseñado en torno a un asistente virtual, capaz de reconocer distintos comandos.

A continuación están expuestas las diferentes funciones con las que cuenta el programa y un resumen de cómo funcionan:

Reconocimiento de voz y habla

El reconocimiento de voz del programa funciona con la API de Google, que puede convertir los sonidos en texto con pequeña latencia. Después las respuestas que genere el código son mandadas al audífono mediante el

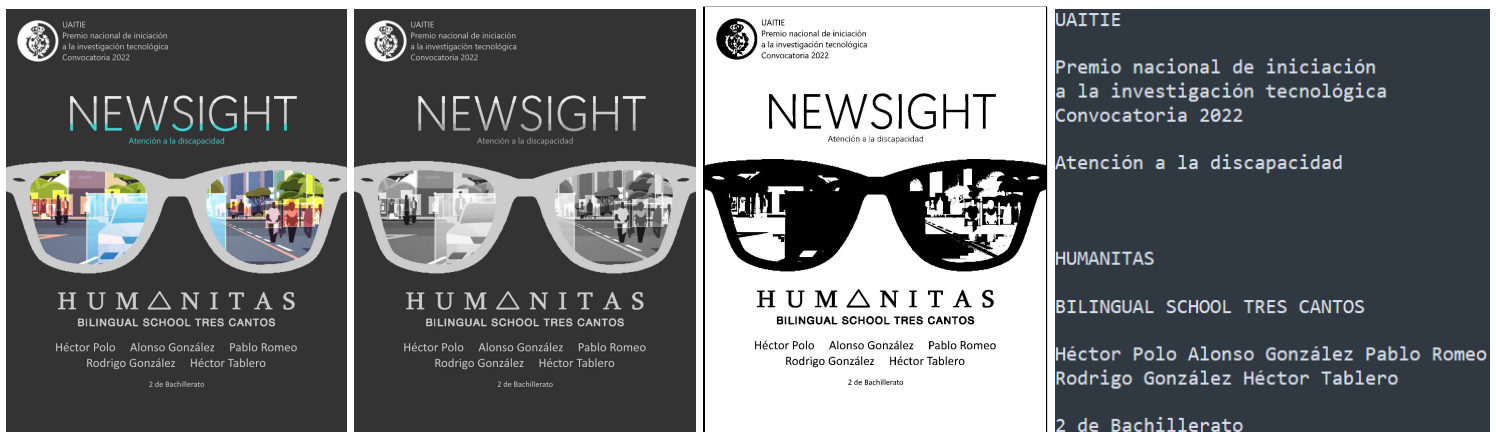
motor pyttsx3, que se encuentra entre los mejores de la plataforma, además de soportar múltiples idiomas.

Búsqueda de información

Usando la interfaz de programación de Wolfram Alpha, Newsight califica lo expresado por el usuario en diversas categorías, y hace una solicitud a la aplicación si determina que se ha hecho una pregunta de temática general, pudiendo ser esta “¿de qué color es x?”, “¿qué tiempo hace en x?” o “¿qué hora es?”.

Lectura de texto

Al recibir este comando, las gafas toman una foto de lo que está viendo en ese momento, y trata de localizar texto en la imagen. De haberlo, preprocesará la imagen utilizando el módulo cv2 de OpenCV, la mayor librería de reconocimiento de imagen en la actualidad. Dependiendo de la claridad de la imagen, puede aumentar la saturación para destacar más el texto o tratar de hacerla más nítida. Una vez la imagen es óptima para su lectura, se reconoce la imagen usando PyTesseract.



Proceso del progreso de edición (foto | escala de grises | blanco y negro con supresión de ruido | resultados)

*En este caso la supresión de ruido se aplica únicamente en “Newsight”. Como se puede observar, en la primera imagen hay tres colores (blanco, gris y azul) y en la tercera lo detecta todo como una misma palabra.

Adicionalmente, el programa hace uso del módulo de traducción para detectar automáticamente el idioma del texto. Si la imagen se ha analizado

en inglés y está en español o viceversa, se leerá de nuevo con el idioma detectado. Esto no solo ayuda a mejorar la precisión del análisis, sino que también puede corregir faltas de ortografía.

Calificación de imágenes

El programa original estaba diseñado en torno a InceptionV3, pero varios inconvenientes nos llevaron a descartarlo. En primer lugar, cargar una red neuronal directamente desde el dispositivo es un proceso muy lento (desde un ordenador portátil puede tardar unos 10 minutos). Además, durante su ejecución hace muchas operaciones, que pueden agotar rápidamente la batería de los móviles. Por ello, llegamos a la conclusión de que la solución más óptima era integrar un programa de computación en la nube, como es Microsoft Azure.

Azure tiene una amplia gama de APIs de visión artificial, por lo que es perfecto para esta demostración. Las que están integradas incluyen describir lo que está viendo el usuario, detectar objetos y marcas en la imagen y encontrar personas.

Traducción

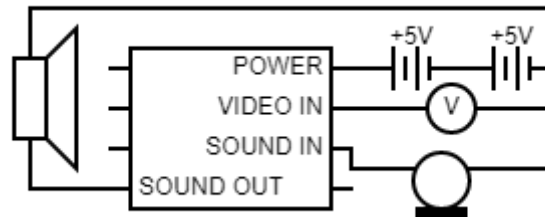
Newsight hace uso de la API de DeepL para traducir algunos de los resultados de los procesos (como los de Wolfram o Azure) al español. Esta función es llamada justo antes de la salida de audio, excepto en el caso del buscador, para el que hace falta traducir las preguntas al inglés primero.

Uno de los requisitos para el traductor de DeepL es indicar el idioma de entrada. Para ello se usa otro traductor: el de Google. Tiene una función de detección de idioma y también puede traducir por cuenta propia. Sin embargo, debido a la menor calidad de los resultados en comparación al traductor principal, solo se usa en caso de fallar el primero.

Etapa de maquetación

A la hora de diseñar las gafas inicialmente consideramos la opción de incorporar un procesador en ellas, pero, en vista de los recursos requeridos y

el límite de espacio, decidimos usar una pequeña placa bluetooth conectada al móvil del usuario, que se encargará de ejecutar el código. Esta estará conectada a un micrófono, a la cámara y al audífono. En cuanto a la alimentación, para mantener la simetría y el balance de peso, habrá una batería en cada patilla. Están conectadas en paralelo y se cargarán únicamente por el lado contrario al de la placa, por conveniencia.



Diseño del circuito

Para la demostración de Newsight en lugar de hacer una maqueta rudimentaria usaremos el ordenador ya que cuenta con altavoz, cámara, micrófono y procesador.

OBJETIVOS DE NEWSIGHT

El objetivo planteado en Newsight es reducir notoriamente las dificultades de tránsito de las personas con problemas visuales. Buscamos que dichas personas con pérdida de visión, parcial o total, logren acercarse lo máximo al entorno sin su problemática. Además, Newsight no sólo será útil para la vía pública sino también para cualquier circunstancia dada en la vida del individuo.

Por otro lado, esperamos que Newsight sea un producto común entre la sociedad para brindar la máxima ayuda posible y que resulte fiable. Por ello, el dispositivo está pensado para que el precio sea asequible y pueda llegar a los clientes sin suponer una carga económica. Sin embargo, la idea del artículo no pretende suplantar los métodos convencionales de orientación y movilidad.

Otro de nuestros objetivos es que nuestra propuesta consiga mermar los accidentes que puedan involucrar a personas invidentes a través de este sistema de guía innovador. Reducir estos acontecimientos nos parece prioritario, y es algo en lo que nuestro producto podrá participar.

CONCLUSIONES

Nuestro proyecto supondría una gran mejora a las condiciones de vida de las personas afectadas por esta discapacidad visual. Cubre la mayoría de las necesidades y está formulado para satisfacer al cliente en la pluralidad de situaciones en las que se puede encontrar.

Está planteado para que funcione en conjunto con las herramientas ya existentes tales como el bastón o los perros guía. No queremos que sea un sustituto de lo que ya es de gran ayuda para los invidentes.

Newsight abre muchas posibilidades. Como ya mencionamos está ideado como boceto. Esto es debido a las limitaciones que impiden alcanzar el potencial real del producto, tales como; el capital, el tiempo y los materiales.

En un futuro podríamos optimizar:

- El tamaño de los componentes de las gafas, para reducir el coste o habilitar la posibilidad de usar distintos materiales
- La calidad de los elementos
- La comodidad
- El precio de producción, para que sea un producto accesible para todo el mundo sin importar su poder adquisitivo
- Las funciones que sería capaz de llevar a cabo las gafas

BIBLIOGRAFÍA

A. D. D. O. C. de España (2015, 10 julio). «La discapacidad no tiene fronteras» DISCAPACIDAD VISUAL D.O.C.E. (DISCAPACITADOS OTROS CIEGOS DE ESPAÑA).

<https://asociaciondoce.com/2015/07/10/la-discapacidad-no-tiene-fronteras-testimoni-o-de-estudiantes-con-discapacidad-visual-en-la-universidad-de-barcelona-texto-integro-de-expansion-com/>

How to classify images with TensorFlow using Google Cloud Machine Learning and Cloud Dataflow. (s. f.). Google Cloud Blog.

<https://cloud.google.com/blog/products/gcp/how-to-classify-images-with-tensorflow-using-google-cloud-machine-learning-and-cloud-dataflow>

Image captioning with visual attention | TensorFlow Core. (s. f.). TensorFlow.

https://www.tensorflow.org/tutorials/text/image_captioning

inception_v3 keras implementation. (s. f.). Github.

<https://gist.github.com/RahulGupta-Reflektion/93b268219ebc3fde7475928b646c71b3>

Keras Classifier Inceptionv3. (2020, 1 septiembre). YouTube.

<https://www.youtube.com/watch?v=CRfBegilH-Y>

Languages supported in different versions of Tesseract. (s. f.). Tesseract.

<https://tesseract-ocr.github.io/tessdoc/Data-Files-in-different-versions.html>

Module: tf.keras.applications.inception_v3 | TensorFlow Core v2.8.0. (s. f.).

TensorFlow.

https://www.tensorflow.org/api_docs/python/tf/keras/applications/inception_v3

Murray, C. (2018, 16 junio). Deep Learning with Keras on Google Compute Engine - Google Cloud - Community. Medium.

<https://medium.com/google-cloud/keras-inception-v3-on-google-compute-engine-a54918b0058>

pytesseract. (2022, 19 febrero). PyPI. <https://pypi.org/project/pytesseract/>

Show and Tell: image captioning open sourced in TensorFlow. (2016, 22 septiembre). Google AI Blog.

<https://ai.googleblog.com/2016/09/show-and-tell-image-captioning-open.html>

K Team. (s. f.-a). Keras documentation: Computer Vision. Keras.

<https://keras.io/examples/vision/>

K Team. (s. f.-b). Keras documentation: Image Captioning. Keras.

https://keras.io/examples/vision/image_captioning/

K Team. (s. f.-c). Keras documentation: InceptionV3. Keras.

<https://keras.io/api/applications/inceptionv3/#inceptionv3>

K Team. (s. f.-d). Keras documentation: Natural Language Processing. Keras.

<https://keras.io/examples/nlp/>

K Team. (s. f.-e). Keras documentation: The Model class. Keras.

<https://keras.io/api/models/model/#model-class>

Wolfram|Alpha: Making the world's knowledge computable. (s. f.). Wolfram Alpha.

<https://www.wolframalpha.com/>

GitHub - ptrstn/deepl-translate: An unofficial python package to translate text using deepl.com. (s. f.). GitHub. <https://github.com/ptrstn/deepl-translate>

googletrans. (2020, 14 junio). PyPI. <https://pypi.org/project/googletrans/>

Detección de objetos: Computer Vision - Azure Cognitive Services. (2022, 25 marzo). Microsoft Docs.

<https://docs.microsoft.com/es-es/azure/cognitive-services/computer-vision/concept-object-detection>

Quickstart: Image Analysis client library or REST API - Azure Cognitive Services. (2021, 18 diciembre). Microsoft Docs.

<https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/quickstart-sdks/image-analysis-client-library?pivots=programming-language-python&tabs=visual-studio>

Image descriptions - Computer Vision - Azure Cognitive Services. (2022, 6 enero).

Microsoft Docs.

<https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/concept-describing-images>

Inteligencia artificial en Azure. (s. f.). Microsoft Azure.

<https://azure.microsoft.com/es-es/get-started/ai/?subscriptionId=dbce46ef-b597-4471-bbee-3bb3e56ea537>