



SISTEMAS INTELIGENTES AUTÓNOMOS PARA EJECUTAR MISIONES DE BÚSQUEDA Y RECONOCIMIENTO

Laura María López Moreno, Juan Álvaro Moya Garzón, David Mauricio Valoyes Porras, Fran Ernesto Romero Álvarez

**Universidad El Bosque
Bogotá, Colombia**

Resumen

Este trabajo presenta el desarrollo de un prototipo de vehículo inteligente autónomo con capacidades para ejecutar sencillas misiones de búsqueda de objetos y reconocimiento de su entorno. El artefacto integra múltiples tecnologías de hardware y software: una tarjeta Raspberry Pi® 4 permite controlar los subsistemas de visión y movimiento autónomo. El acelerador USB Google Coral® permite ejecutar con mayor velocidad diversos algoritmos de inteligencia artificial, mejorando considerablemente el desempeño de la tarjeta Raspberry. En cuanto al software, se ha adoptado Python 3® como lenguaje de programación junto con la librería TensorFlow®, la cual permite la implementación de algoritmos como *YOLO -You Only Look Once-* usado para el reconocimiento de objetos en tiempo real y el algoritmo *SLAM -Simultaneous Localization And Mapping-* el cual permite que el vehículo explore su entorno de forma autónoma. Las aplicaciones prácticas del vehículo varían en un amplio rango, desde servir como juguete didáctico hasta explorar ambientes inaccesibles o inhóspitos para un ser humano.

Desde el punto de vista académico, el desarrollo de un Sistema Inteligente Autónomo constituye un claro ejemplo de cómo la educación STEM -*Science, Technology, Engineering and Math-* integrada con el modelo pedagógico de la Universidad El Bosque, conocido como Aprendizaje Significativo, permite potenciar las capacidades de los estudiantes a nivel individual y en equipo, a la vez que se estimula la innovación y el emprendimiento de base tecnológica. Se espera que los futuros profesionales tengan los conocimientos y habilidades para transformar positivamente los hábitos y creencias del entorno social y productivo a través del desarrollo y la transferencia efectiva de artefactos tecnológicos basados en Inteligencia Artificial.

Palabras clave: sistema inteligente autónomo; visión artificial; movimiento autónomo

Abstract

This work presents the development of a prototype of an autonomous intelligent vehicle with the capabilities to carry out simple missions of searching for objects and recognizing its surroundings. The device integrates multiple hardware and software technologies: a Raspberry Pi® 4 card allows control of the vision and autonomous movement subsystems. The Google Coral® USB accelerator allows various artificial intelligence algorithms to be executed faster, considerably improving the performance of the Raspberry card. As for the software, Python 3® has been adopted as a programming language together with the TensorFlow® library, which allows the implementation of algorithms such as YOLO -You Only Look Once- used for real-time object recognition and the SLAM algorithm. -Simultaneous Localization and Mapping- which allows the vehicle to explore its surroundings autonomously. The practical applications of the vehicle vary in a wide range, from serving as a teaching toy to exploring environments that are inaccessible or inhospitable for a human being.

From the academic point of view, the development of Intelligent Autonomous Systems constitutes a clear example of how STEM education -Science, Technology, Engineering and Math- integrated with the pedagogical model of El Bosque University, known as Meaningful Learning, allows to enhance the capacities of students individually and as a team, while stimulating innovation and technology-based entrepreneurship. Future professionals are expected to have the knowledge and skills to positively transform the habits and beliefs of the social and productive environment through the development and effective transfer of technological artifacts based on Artificial Intelligence.

Keywords: *autonomous intelligent system; artificial vision; autonomous movement*

1. Introducción

Los Sistemas Inteligentes Autónomos - SIA utilizan Inteligencia Artificial – IA para aprender de su entorno y planificar sus acciones a partir de las reacciones obtenidas del medio en el cual se desenvuelven. Un SIA puede ser robot virtual –bot- que entabla conversaciones con seres humanos o con otros bots, aprende de estas conversaciones y ejecuta acciones al respecto, tal como realizar una inversión financiera o contrarrestar un ciberataque. Un robot físico también es un SIA, ya que utiliza sensores para captar información de su medio, aprende de dicha información y toma decisiones, tal como un dron diseñado para entregar paquetes o un vehículo terrestre autónomo. El desarrollo de SIA supone la integración de diversas tecnologías, como robótica, inteligencia artificial, aprendizaje de máquina, sistemas embebidos, sensórica, desarrollo de aplicaciones móviles, visión artificial y Procesamiento de Lenguaje Natural – PLN.

El Semillero de Investigación en SIA de la Universidad El Bosque, ha venido trabajando en diversos proyectos enfocados en atender necesidades en diversas comunidades: bots conversacionales para psico-educación en la población infantil y drones inteligentes autónomos como herramienta para el



control del distanciamiento social a propósito de la pandemia del COVID-19. Recientemente el equipo ha venido trabajando en el desarrollo de un prototipo de vehículo terrestre con capacidades de visión artificial y navegación autónoma. El vehículo está en capacidad de reconocer su entorno, evadir obstáculos y reconocer diversos objetos, pudiendo buscar y localizar uno en específico. Las aplicaciones de esta tecnología varían en un amplio rango, sirviendo como juguete didáctico en entornos de aprendizaje, exploración de ambientes inhóspitos o de difícil acceso y control de cultivos, solo para citar algunos ejemplos. En las siguientes fases del proyecto, se espera integrar al SIA con funcionalidades para reconocimiento de comandos de voz y capacidades de computación afectiva, siendo capaz de reconocer y reaccionar a diversas emociones humanas.

2. Arquitectura del SIA

El vehículo inteligente está gobernado por una tarjeta *Raspberry Pi® 4*, la cual permite ejecutar los diferentes programas relacionados con visión artificial y movimiento autónomo. Se evaluaron diferentes modelos de tarjetas programables, tales como *NVIDIA Jetson Nano®*, *Odroid N2* y *Banana Pi M5®*, las cuales ofrecen características similares; finalmente se decidió optar por *Raspberry Pi 4®*, debido a su disponibilidad en el mercado y mejor relación costo-beneficio. Esta tarjeta posee un procesador ARM Cortex-A72 y permite trabajar con video 4K a 60fps, siendo ideal para aplicaciones que requieran procesamiento de video. Respecto al sistema operativo, se utiliza *Raspberry Pi OS*, el cual está basado en la distribución Debian de Linux y permite ejecutar sin problemas programas en Python con uso de múltiples librerías para aprendizaje automático - TensorFlow- y procesamiento de imágenes -OpenCV-.

El sistema utiliza diversos algoritmos de inteligencia artificial para llevar a cabo las tareas de navegación, búsqueda y reconocimiento del entorno (y a futuro de ejecución de órdenes a través de comandos de voz). La ejecución de estos algoritmos conlleva a realizar una gran cantidad de operaciones de punto flotante por segundo, siendo necesaria una alta capacidad de procesamiento que la tarjeta principal no está en capacidad de proveer. Para solventar esta problemática, ha sido necesario agregar un componente *Coral® USB Accelerator*, el cual permite acelerar los cálculos de inferencia requeridos para los algoritmos de aprendizaje automático. A través de su coprocesador integrado *Edge TPU*, este dispositivo puede realizar 4 trillones de operaciones por segundo, usando únicamente 2W de potencia. Otra característica importante es la alta compatibilidad con los modelos generados a partir de TensorFlow®, los cuales pueden ser fácilmente compilados para su uso con *Coral*.

La captura de video se realiza a través de una cámara Mini Vision OV564, la cual provee una resolución de 5 megapíxeles y trabaja a una velocidad de 30 cuadros por segundo, suficientes para los requerimientos del vehículo. Las imágenes son inicialmente procesadas utilizando la librería OpenCV y posteriormente son enviadas a un modelo de TensorFlow Lite®, el cual se ejecuta en el dispositivo *Coral* para reconocer diferentes objetos presentes en la escena. A partir de las posiciones de los diversos objetos detectados, el vehículo puede determinar distancias y ángulos que le permiten ajustar su rumbo y ayudar a conocer su posición en el entorno.



Para controlar los motores ha sido necesario utilizar el *driver* L298N, el cual permite manejar simultáneamente 2 motores DC de hasta 2 amperios, controlando el sentido y velocidad de giro mediante señales TTL proveniente de la tarjeta Raspberry. Específicamente, la velocidad de giro se puede regular haciendo uso de modulación por ancho de pulso – PWM. Por otra parte, un sensor de ultrasonido HC-SR04 instalado en la parte frontal del vehículo evita que el mismo colisione con objetos a una distancia menor de 20cm; para su funcionamiento, este sensor envía un pulso de arranque y mide la anchura del pulso de retorno, siendo capaz de detectar objetos y calcular la distancia a la que se encuentran en un rango de 2cm a 400cm.

Finalmente, para la alimentación eléctrica del artefacto se utilizan 2 baterías separadas: una batería LiPo de 1500mA se encarga de proveer energía para los 4 motores DC y un *Power Bank* provee alimentación necesaria para la tarjeta Raspberry, la cual a su vez alimenta el dispositivo *Coral USB* y el sensor de ultrasonido. Se optó por este sistema de alimentación independiente para prevenir daños por corto circuito en la tarjeta principal y demás dispositivos conectados. La figura 1 muestra la arquitectura general del sistema.

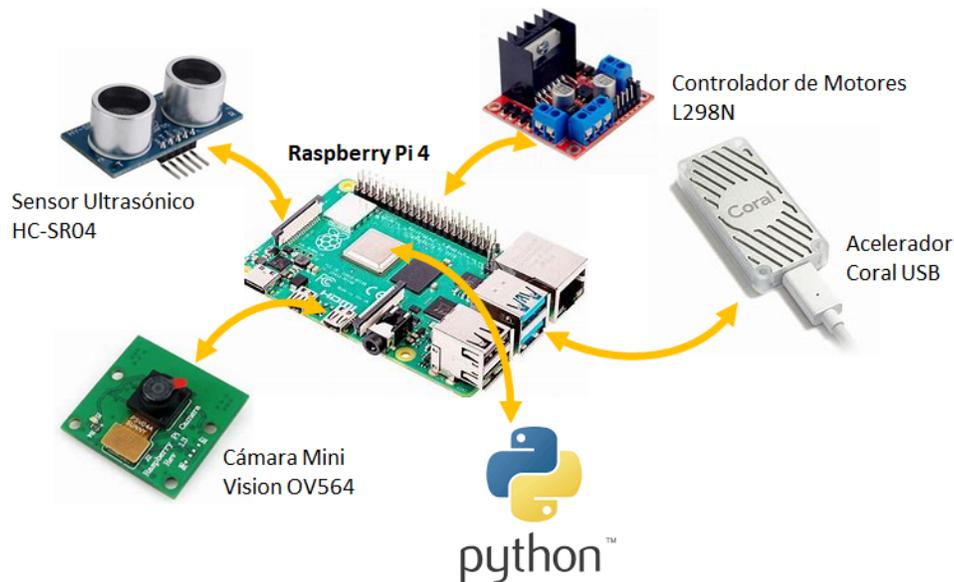


Figura 1. Componentes del Vehículo Inteligente Autónomo

3. Vision artificial

La detección de objetos constituye un aspecto clave en el ámbito de los Sistemas Inteligentes Autónomos, tomando en cuenta que uno de sus objetivos es poder percibir y comprender su entorno. Los enfoques clásicos de Visión Artificial basados en Aprendizaje Automático, como aquellos que utilizan Máquinas de Soporte Vectorial para tareas de clasificación, tienden a tener un tiempo de respuesta lento. Con el fin de mejorar la eficiencia en el reconocimiento de objetos en imágenes de video, existen algoritmos modernos basados en técnicas de Aprendizaje Profundo. En el caso del vehículo autónomo, se ha implementado el algoritmo *YOLO (You Only Look Once)*, el cual utiliza Redes Neuronales Convolucionales – RNC para identificar objetos en tiempo real y sin incurrir en pérdidas significativas de precisión.



Para lograr su objetivo, YOLO divide cada imagen muestreada del video en un conjunto de subcuadrículas, a cada una de las cuales se asigna una probabilidad de representar un objeto específico. Una RNC es entrenada para determinar la probabilidad de que una subcuadrícula - caracterizada por su centro, alto y ancho- contenga un objeto de determinada clase. Las capas iniciales de la red extraen las características de la imagen, mientras que las capas completamente conectadas predicen las probabilidades y coordenadas de salida. La figura 2 ilustra este proceso.

A diferencia de otros sistemas de detección, el algoritmo YOLO replantea la detección de objetos como un solo problema de regresión, directamente desde los píxeles de la imagen hasta las coordenadas del cuadro delimitador. El modelo entrenado solo "observa" una vez una imagen para predecir que objetos están presentes y donde están ubicados. Como se puede observar en la figura 1, una sola red convolucional predice simultáneamente múltiples cuadros delimitadores y las probabilidades de clase para cada uno de esos cuadros.

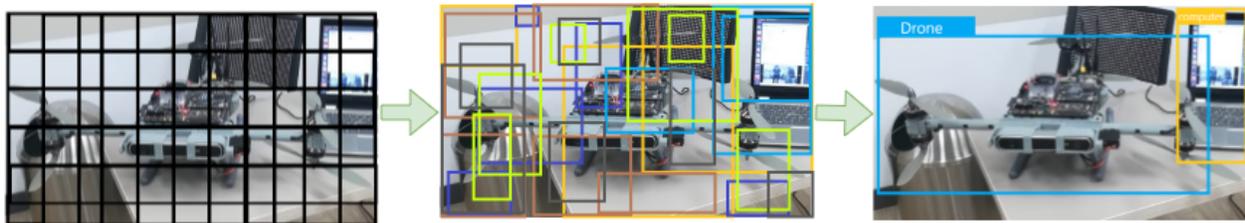


Figura 2. Algoritmo YOLO

Este modelo unificado presenta varios beneficios sobre los métodos tradicionales de detección de objetos. En primer lugar, *YOLO* es bastante rápido: desde el momento de la detección de cuadros como problema de regresión, no se necesita un proceso de transformación complejo de la imagen y la RNC identifica en una sola pasada los cuadros y las probabilidades de cada región. En el caso específico del vehículo, el algoritmo puede analizar imágenes a una velocidad de 4fps utilizando únicamente los recursos de procesamiento de la tarjeta Raspberry; utilizando el dispositivo *Coral USB* la velocidad se incrementa fácilmente a 15fps.

Otra ventaja de *YOLO* es que aprende representaciones generales de los objetos, siendo menos probable que falle cuando se encuentren entradas inesperadas. Esta característica los hace muy conveniente para la exploración de espacios desconocidos, aunque por otra parte tenga dificultades para localizar con precisión objetos pequeños.

La figura 3 ilustra el proceso de reconocimiento de objetos en el vehículo inteligente autónomo. En primer lugar, el video es capturado por la cámara *Mini Vision*. Del *stream* de video, se procesan 15 imágenes por segundo, cada una de las cuales es procesada con ayuda de la librería *OpenCV* y el resultado es enviado al modelo de reconocimientos de objetos -el cual implementa *YOLO*- en *TensorFlow Lite*. Como se mencionó antes, el dispositivo *Coral USB* acelera los cálculos requeridos para el reconocimiento de imágenes.



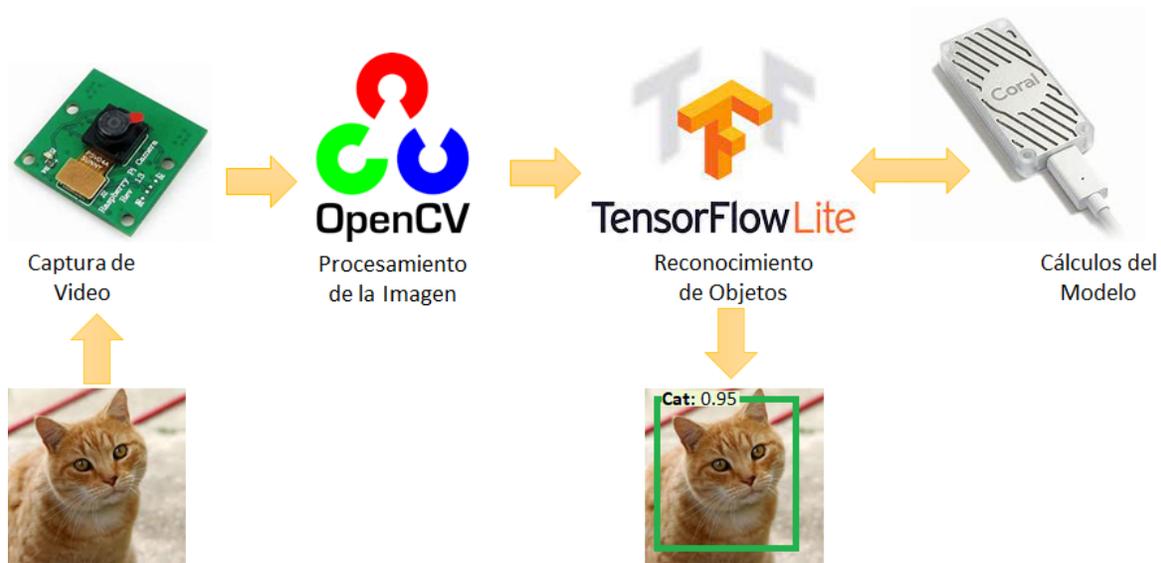


Figura 3. Reconocimiento de imágenes en el vehículo autónomo

4. Reconocimiento de voz

En la actualidad, la interacción entre humano y máquina a través de la voz ha venido aumentando de forma acelerada. Los sistemas de reconocimiento de voz generalmente corresponden a interfaces de programación disponibles en la nube, tales como *Apple Siri®*, *Google Assistant®*, *Microsoft Cortana®* y *Amazon Alexa®*. En el caso del vehículo autónomo, se ha desarrollado un sistema de reconocimiento de voz que no depende de una conexión a Internet, aplicando el concepto de computación al borde de la red (*Edge Computing*). Este sistema permite reconocer una cantidad limitada de comandos de voz en español, logrando obtener un entorno más seguro y con menores tiempos de respuesta.

Los modelos basados en RNC también suelen ser utilizados para tareas de reconocimiento de voz. En este caso, las señales de audio son convertidas en espectrogramas, los cuales corresponden a mapas de calor que representan los cambios de frecuencia y amplitud a través del tiempo en formato de imagen de 2D. Las imágenes generadas son posteriormente utilizadas para entrenar una RNC que permita la clasificación del audio, logrando excelentes tasas de efectividad. La figura 4 muestra la señal y el espectrograma asociados a la palabra "Adelante".



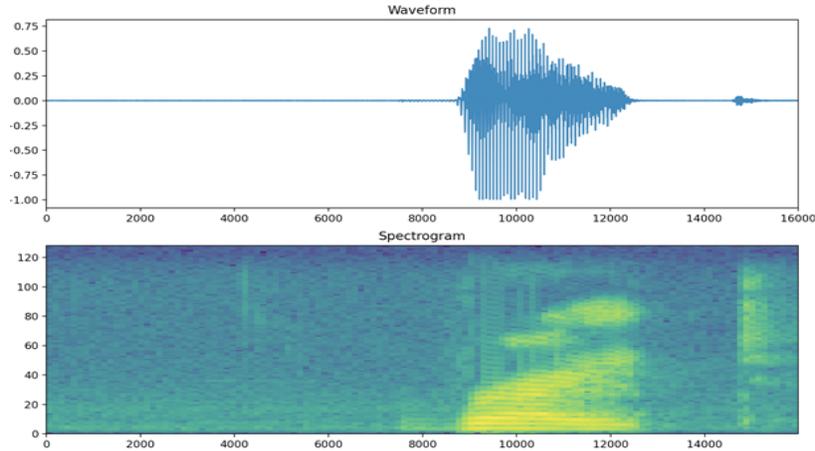


Figura 4. Espectrograma de una señal de audio

A fin de obtener un conjunto de entrenamiento suficientemente grande, se solicitó a un grupo de 50 personas que realizaran grabaciones de 1 segundo para cada uno de los comandos de voz objetivo; estas muestras no fueron suficientes para alcanzar un nivel de exactitud adecuado, por lo que ha sido necesario utilizar una técnica de aumentación de datos. Este proceso consiste en generar nuevas muestras a partir de las originales, para lo cual se modifican ciertos parámetros asociados al ruido, el volumen y el tono. La aumentación de datos permitió incrementar 4 veces el tamaño del conjunto de entrenamiento.

Finalmente, se procedió a entrenar un modelo de RNC en TensorFlow. Para este fin, se usó el 80% de las muestras para el entrenamiento, el 10% para validación y el 10% restante para pruebas. Finalizado este proceso, se obtuvo un 70% de exactitud, lo que indica que se debe entrenar el modelo con más muestras. Una vez el modelo sea adecuadamente calibrado, se compilará para ser utilizado con el dispositivo *Coral USB*, a fin de optimizar los tiempos de respuesta.

5. Movimiento autónomo

Las capacidades de navegación autónoma constituyen uno de los retos más importantes en el diseño del vehículo inteligente. Un SIA debe tener la capacidad para reconocer el entorno en el cual se encuentra operando, conocer la localización de sí mismo dentro del entorno y evitar o rodear diversos obstáculos, tales como objetos, paredes y accidentes del terreno. Se han considerado diversas alternativas para abordar el problema del movimiento: en primer lugar, se implementaron las funcionalidades necesarias para hacer que el vehículo siga una ruta predefinida, de tal forma que, al principio de una misión de exploración, el artefacto cuenta con un mapa y un conjunto predefinido de coordenadas específicas que debe visitar. Esta aproximación presentó diversos inconvenientes: primero, supone un conocimiento previo del entorno, lo cual no siempre es posible; por otra parte, resulta difícil programar que el artefacto avance y gire con precisión, pues ello depende de múltiples variables como las condiciones del terreno, ángulos de pendiente, velocidad y energía de la batería.



Una segunda aproximación al problema de la navegación autónoma es el uso de sensores de proximidad. Esencialmente, el vehículo lleva a cabo una “caminata aleatoria” y gracias a los sensores es posible evitar o rodear obstáculos. Las desventajas de este enfoque radican en que el SIA puede quedarse dando vueltas en un mismo lugar y realmente no tiene forma de conocer su posición en el entorno. A continuación, se muestra el algoritmo usado para esta estrategia:

ALGORITMO *Caminata Aleatoria*

$N = 30$

$G = 90$

while *true* **do**

$d = \text{medir_espacio_libre_frontal}()$

if $d < N$ **then**

$\text{girar}(G)$ //Gira G -grados

end if

$\text{avanzar}(N)$ //Avanza N -centímetros

end do

Finalmente, se ha decidido llevar a cabo la implementación del algoritmo SLAM -Localización y Mapeo Simultáneos-. SLAM es un método que intenta construir un mapa del entorno desconocido, mientras usa el mismo mapa para determinar la ubicación de un robot dentro del mapa. Este algoritmo se puede utilizar para limitar el error cuando se realiza una navegación sin instrumentación externa y, por lo tanto, se considera que SLAM es un requisito para operaciones verdaderamente autónomas. En realidad, SLAM corresponde a una familia de algoritmos cuya implementación depende del tipo de sensor con el que se cuente y el tipo de aproximación matemática utilizada, siendo las de tipo probabilístico los que mejores resultados han mostrado.

Gracias a que el vehículo cuenta con capacidades de visión artificial, es posible implementar la versión de SLAM “basada en características”, en la cual se almacena un conjunto de puntos de referencia a partir de las lecturas de un sensor, tales como un láser, un sonar o, como en este caso, una cámara. Los enfoques sin características a menudo se prefieren en entornos donde las características claramente identificables son escasas. En este momento se trabaja en optimizar la técnica para los recursos de procesamiento limitados de la tarjeta Raspberry Pi.

6. Conclusiones

La integración de diversas tecnologías de visión artificial, movimiento autónomo y reconocimiento de voz en un Sistema Inteligente ha conllevado superar diferentes desafíos de ingeniería, aún más si se toma en cuenta que el artefacto ha sido desarrollado sobre una tarjeta con capacidades de procesamiento limitadas como Raspberry Pi®. Continuar con la investigación y desarrollo de este tipo de tecnologías, las cuales integran robótica e inteligencia artificial, constituye una necesidad imperiosa para el crecimiento y competitividad de diversos sectores productivos. Conviene que las facultades de ingeniería preparen adecuadamente al capital humano que habrá de poner este conocimiento al servicio del país.



El futuro se espera perfeccionar el artefacto tecnológico con miras a encontrar aplicaciones específicas en sectores tales como la educación, el comercio o el turismo; las posibilidades son infinitas. El segundo eje de trabajo será mejorar progresivamente la interacción por voz, de tal forma que el robot pueda recibir instrucciones más complejas y llevarlas a cabo. A largo plazo, el equipo de trabajo espera llegar a incorporar capacidades de *Computación Afectiva*, permitiendo que el robot pueda reaccionar a diversas emociones humanas.

El desarrollo de este proyecto constituye un ejemplo claro de cómo la aplicación de la metodología STEM fomenta, no sólo el aprendizaje de ciencias, tecnología, ingeniería y matemáticas, también el trabajo en equipo y la habilidad para desarrollar proyectos complejos. El prototipo ha sido construido al interior de un semillero de estudiantes de Ingeniería de Sistemas, quienes han tenido que investigar y apropiarse temas de ingeniería electrónica y otras disciplinas, demostrándose a sí mismos que cuando se trabaja de manera metodológica y disciplinada, siempre será posible alcanzar los objetivos propuestos.

7. Referencias

Artículos de revistas

- P. Patel, A. S. A. Doss, L. PavanKalyan, and P. J. Tarwadi. (2021). Speech Recognition Using Neural Network for Mobile Robot Navigation. *Trends in Mechanical and Biomedical Design*, pp. 665–676.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788.
- Durrant-Whyte, H., & Bailey, T. (2006). Simultaneous localization and mapping: part I. *IEEE Robotics & Automation Magazine*, 13(2), pp. 99–110. doi:10.1109/mra.2006.1638022.
- M. Stohlmann, T. J. Moore, and G. H. Roehrig. (2012). Considerations for teaching integrated STEM education. *Journal of Pre-College Engineering Education Research (J-PEER)*, Vol. 2, No. 1, pp. 28–34.

Sobre los autores

- **Laura María López Moreno**. Estudiante de Ingeniería de Sistemas - Universidad El Bosque. lmlopezm@unbosque.edu.co
- **Juan Alvaro Moya Garzon**. Estudiante de Ingeniería de Sistemas – Universidad El Bosque. jmoyag@unbosque.edu.co
- **David Valoyes Porras**. Estudiante de Ingeniería de Sistemas – Universidad El Bosque. dvaloyesp@unbosque.edu.co
- **Fran Ernesto Romero**. Profesor asistente del programa de Ingeniería de Sistemas – Universidad El Bosque. fromeroa@unbosque.edu.co



Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2021 Asociación Colombiana de Facultades de Ingeniería (ACOFI)

