

# Implementación de sistema de visión inteligente para reconocimiento de rostros en robot de código abierto BOB

Carlos Leal-Beltrán<sup>2</sup>, Juan Villegas-Cortez<sup>1</sup>, Arturo Zúñiga-López<sup>2</sup>,  
Salomón Cordero-Sánchez<sup>3</sup>

<sup>1</sup> Universidad Autónoma Metropolitana,  
Departamento de Sistemas, Unidad Azcapotzalco,  
México

<sup>2</sup> Universidad Autónoma Metropolitana,  
Departamento de Electrónica, Unidad Azcapotzalco,  
México

<sup>3</sup> Universidad Autónoma Metropolitana,  
Departamento de Química, Unidad Iztapalapa,  
México

car-lebe@hotmail.com, {juanvc, azl}@azc.uam.mx,  
scs@xanum.uam.mx

**Resumen.** En el presente trabajo se realiza la implementación de una interfaz entre un mecanismo móvil de código abierto y un sistema de autenticación de rostros para personas, ya que el diseño recuerda mucho a un juguete lo vuelve más amigable con niños y jóvenes que otro tipo de mecanismos y robots, lo que lo vuelve más accesible y en consecuencia hace más fácil la interacción humano-computadora, pudiendo interfazarse con otros proyectos de evaluaciones pedagógicas, de educación o con otros sistemas de reconocimiento de personas, independientemente del lenguaje o propósitos de comunicación. Proponemos utilizar la interconexión, con la posibilidad de reconfiguración y mejora para facilitar la comunicación entre personas, a partir de recabar la información de rostros. El sistema consta de una tarjeta de sistema embebido Raspberry Pi Zero W, una cámara tipo CSI, y la implementación de una red neuronal embebida de código abierto también. Se ha realizado el ensamblado del mecanismo y la integración del sistema, logrando una operación fiable y las pruebas de reconocimiento están basadas en dos bases de datos de rostros probadas por la comunidad internacional de inteligencia artificial.

**Palabras clave:** Robótica, visión por computadora, redes neuronales profundas, reconocimiento de rostros.

## Smart Computer Vision System Applied to Face Recognition in Open Source Robot BOB

**Abstract.** This work studies the implementation of an interface between an open code mobile mechanism and a person face authentication system.

The design of it is similar to a child's play, which makes it more attractive to children and youths than any other type of mechanism and robot. In this sense, it is more accessible and makes human computer interaction easier. Also, stands out its capability to interface with other pedagogical projects and with other face recognition systems, regardless language or communication purposes. Based on collected face information data, it is proposed to utilize the interconnection with the possibility of continuous reconfiguration and improvement to facilitate communication among people. The system consists of a Raspberry Pi Zero W embedded card system, a CSI camera, and the implementation of a neural network embedded in open code. The assembly of the mechanism and its integration achieve a reliable operation and the recognition tests are based on two face databases tested by the international community of artificial intelligence.

**Keywords.** Robotics, computer vision, deep neural networks, pattern recognition.

## 1. Introducción

La ciencia informática camina actualmente hacia la creación de computadoras cada vez más rápidas, expertas y autónomas. Una de las empresas más ambiciosas consiste en dotar a las computadoras de la facultad de relacionarse con su entorno del mismo modo que lo hace un humano: a través de los sentidos. El proporcionar capacidad sensorial a una computadora es una tarea difícil ya que entran en juego los periféricos al microprocesador y sus tipos de interfase de comunicación, como son: los sensores, tarjetas adaptadoras de señal o el ruido del entorno. A pesar de los muchos inconvenientes, existe hoy un interés especial por dotar a las computadoras de uno de los cinco sentidos del hombre: la habilidad de ver. Es así que la visión artificial se aplica ya en procesos industriales en los que la simple detección de presencia no resulta una fuente de información suficiente [6].

En este trabajo proponemos implementar una metodología general de reconocimiento de rostros en un sistema embebido portable, con programación en código abierto y usando herramientas par lograr su interfase hacia otros sistemas de propósito específico. Aquí nosotros presentamos la propuesta implementada en un mecanismo de código abierto, BOB, y herramientas ampliamente conocidas por la comunidad científica del área de visión por computadora.

El resto del trabajo está organizado de la siguiente forma. En la Sección 2 presentamos el Estado del arte, en la Sección 3 la metodología de detalla con el propósito que pueda ser replicada por la comunidad científica e ingenieril interesada, en la Sección 4 presentamos la implementación en hardware, las pruebas realizadas hasta ahora y una discusión de resultados, y finalmente en la Sección 5 presentamos las conclusiones de nuestro trabajo.

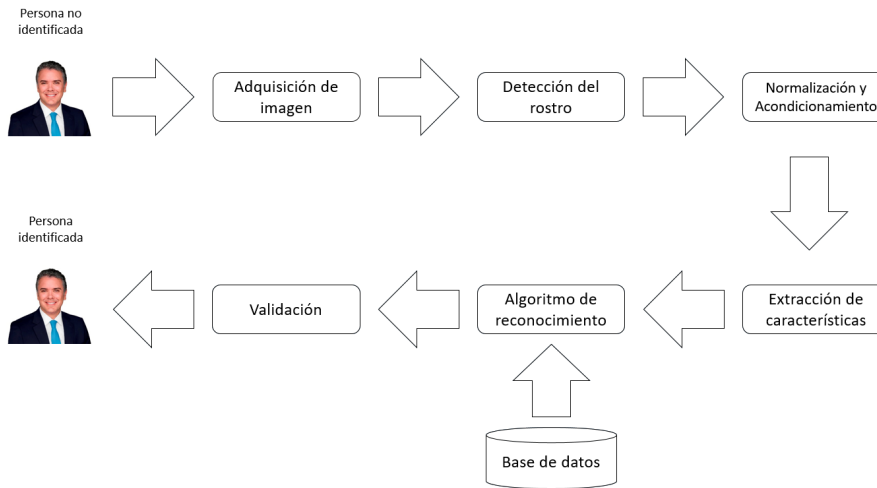


Fig. 1. Diagrama de funcionamiento de un sistema de autenticación de rostros.

## 2. Estado del arte

Los sistemas de visión integrados o sistemas de visión artificial incluyen sistemas de visión, tales como: cámaras inteligentes, tarjetas de adquisición de imágenes e interfaces de visión diseñados para una amplia variedad de aplicaciones como verificación de ensamble, rastreo de objetos y robótica guiada por visión, como puede verse en la página de referencia de la aplicación de visión por computadora en la industria por David Lowe en [5] y también tenemos ejemplos concretos en [3] [9]. En la Figura 1 se muestra un ejemplo de un sistema o metodología general de validación de rostros por reconocimiento facial. Es así que en general, se procesa la imagen de entrada en un sistema que puede categorizar un rostro mediante un clasificador de distancia no supervisado como por ejemplo el de  $K$  vecinos más próximos,  $K - NN$  [2], extrayendo características y medidas de la imagen como un patrón para reconocer rostros por medio de una métrica con base a la cercanía de los patrones de los rostros. Estos datos se cotejan contra una base de datos ya existente previamente entrenada, y en el caso que encuentre quién es la persona de la cual se detectó el rostro por medio de una comparación basada en una métrica de distancia, detecte o identifique a la persona.

El reconocimiento de rostros, del inglés *face recognition* (FR), es una tarea que realizamos las personas de manera cotidiana y después de lo cual toman decisiones y actúan de diferente manera.

Por ejemplo, si se reconoce a una persona conocida se puede proceder a saludarla. Es así que el FR un proceso muy complejo que se define en varias partes, por un lado, la primera de ellas es el momento de tomar la observación o en nuestro caso la foto que se usara, ya que actualmente la mayoría de aplicaciones que usan FR ocupan una ventana de tiempo para decidir en qué

momento realizarán su funcionalidad [13]. Por otro lado, en nuestra propuesta aquí presentada, primero se usó sensores de proximidad, para determinar la ventana de observación en la que se realizará una captura del rostro mediante una cámara; posteriormente, después de la observación y adquisición de la foto de la persona, es el FR en la foto, en este paso también es donde se delimitan los parámetros que podrá tener la imagen como lo son el ángulo, cantidad de luz, uso de lentes y demás parámetros. Algo en lo que en la actualidad se realiza ocupando varias librerías de software para la detección de rostros que realizan el trabajo de normalizar y acondicionar la foto.

En este paso se utiliza una red neuronal artificial, que pueden ser del tipo *eigenfaces*, *fisherface*, etc. [14]. Estos algoritmos de reconocimiento, desde el reconocimiento de patrones, son aquellas instrucciones que están orientadas hacia la agrupación de las características principales en los patrones que nos permitirán distinguir una cara de otra, en este segmento podemos mencionar a los métodos basados en clasificadores bayesianos, redes neuronales artificiales o los algoritmos genéticos y la generalidad de los algoritmos evolutivos [9], que proponen soluciones muy sencillas, desde la generalidad de los Algoritmos Evolutivos, tal como tenemos un ejemplo muy concreto en [11].

El Aprendizaje Profundo, del inglés *Deep Learning* (DL), es una metodología que caracteriza a los objetos de estudio, como puede ser una imagen con todos sus píxeles y capas o canales, en un cúmulo de rasgos obtenidos por etapas sucesivas con un alto costo de cómputo. Desde el enfoque del DL, se considera que sea el propio algoritmo de aplicaciones sucesivas de extracción de características por operaciones de convolución discreta, que se conformen los patrones característicos del objeto de estudio [15]. En esta línea se cuenta con desarrollos abiertos para la comunidad científica en FR aplicando DL, específicamente hablamos de OpenFace <sup>4</sup>, y en nuestra propuesta la retomamos para usarla en el sistema embebido o empotrado basado en Raspberry Pi. En el trabajo de [10] se tiene un ejemplo de este tipo de soluciones, basadas en sistemas embebidos de bajo coste y de redes neuronales profundas; pero ahora nosotros presentamos una implementación orientada a un abanico de posibilidades dentro de la educación y de apoyo a tareas más allá de la autenticación de un usuario o persona concreto. La mayor diferencia o aportación de este proyecto, es que implementa este tipo de soluciones en un mecanismo móvil que interactúa directamente con las personas y sus traducen en respuestas físicas.

### 3. Metodología

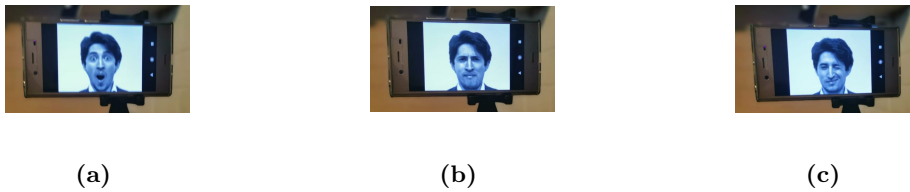
El siguiente paso es el procesamiento de todos estos datos y en el caso del trabajo que se realiza se eligió un dispositivo Raspberry que ya ha probado ser un dispositivo de alto rendimiento que después de pasar un entrenamiento en otro dispositivo de mayor capacidad, es capaz de lograr un reconocimiento facial de gran calidad debido a sus prestaciones que son muy buenas para su tamaño

<sup>4</sup> <https://cmusatyalab.github.io/openface/>

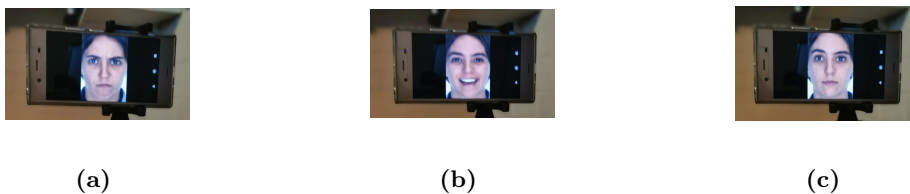


[10]. Como parte final se programó la acción resultante del mecanismo para el movimiento, dependiendo de si el resultado del reconocimiento facial es positivo o en caso contrario tomar otra acción. Para el entrenamiento y pruebas de la red neuronal se optó por las bases de datos de rostros ampliamente conocidas por la comunidad de Visión por Computadora y Reconocimiento de patrones: Yale y Ar, como se muestran en las Figuras 2 y 3 respectivamente. La primera, contiene imágenes de tamaño  $320 \times 243$  píxeles en formato de archivo con compresión JPEG, con 13 individuos, con variaciones de iluminación y con 8 distintas expresiones faciales tales como de postura normal, feliz, triste, soñoliento, etc. La segunda es de fotografías de 83 individuos, todas en condiciones ideales, de frente y con expresiones faciales, y no es regular en la cantidad de muestras por persona para unos tiene 6 poses, mientras otros tienen 8, y también presentan iluminación variante, con resolución de  $576 \times 768$  píxeles.

A continuación damos detalles de la implementación de la metodología en su arquitectura, componentes electrónicos y de sistema usados, así como de las etapas de la metodología que se desarrollaron en una computadora tipo PC, y las etapas desarrolladas en el sistema embebido Raspberry.



**Fig. 2.** Ejemplos de rostros de la base de imágenes Yale tomadas por el mecanismo.



**Fig. 3.** Ejemplos de rostros de la base de imágenes AR captadas por el mecanismo.

## Arquitectura

El sistema que se implementa se muestra en la Figura 4, en un diagrama a bloques con los procesos a realizar. En este sistema se analiza el movimiento y velocidad del mecanismo una vez que se ensamble para verificar que este movimiento no afecte significativamente la eficiencia del experimento.

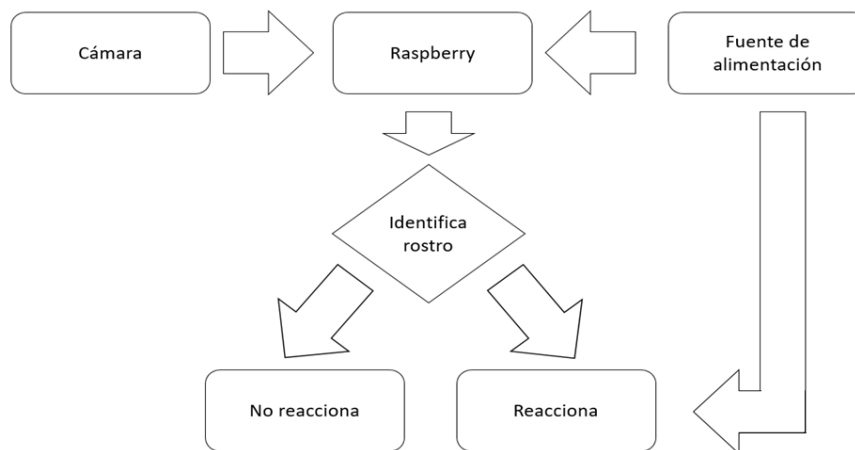


Fig. 4. Diagrama a bloques del funcionamiento del sistema propuesto.

## Componentes

- Sistema embebido: Es el cerebro del proyecto, consiste en un sistema miniaturizado de control que mediante circuitos con interfaces y de procesamiento que se pueden adaptar a cualquier proceso que sea necesario.

Este tipo de sistemas compactos son muy útiles cuando se necesita procesamiento de datos en tiempo real y más usado en aplicaciones móviles. La gran ventaja de estos sistemas es que son programables, lo que lo vuelven en un sistema muy elástico.

En cuanto a programación tenemos muchos lenguajes para programarlo y de diferentes tipos tales como lenguajes ensamblador, lenguajes orientados a objetos, etc. Con lo que tenemos diferentes herramientas para las aplicaciones que necesitemos y del cual se usará un lenguaje orientado a objetos Python. Para aplicaciones con sensores de visión o cámaras útiles para el proyecto propuesto se recomiendan sistemas del tipo *Raspberry Pi Zero W*, que es el más adecuado por sus especificaciones, tales como son memoria, procesador e interfaces como se muestra en la Figura 5.

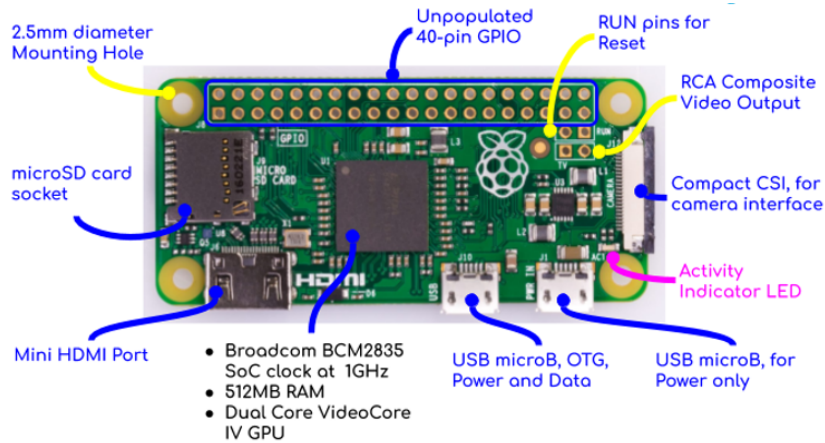


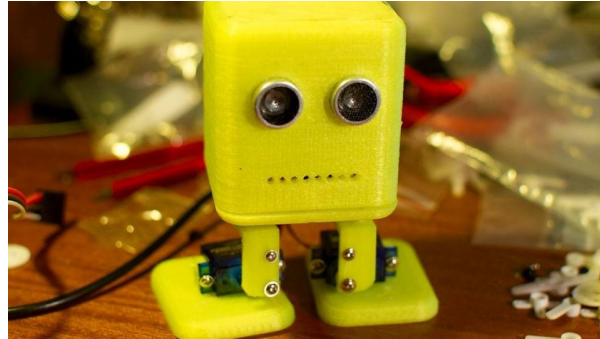
Fig. 5. Componentes de la Raspberry Pi Zero W.

#### – Características del sistema embebido

- Procesador tipo: BCM 2835 SOC.
  - Velocidad de procesador: 1GHz.
  - Memoria RAM: 512MB.
  - Memoria ROM: micro-SD.
  - Puerto de Audio y Video: mini-HDMI.
  - Puerto USB: 1 x micro-B USB para datos (y alimentación también).
  - Entrada de alimentación: 1x micro-B USB for power (sin datos).
  - Interfaz para Cámara: conector CSI para camera.
  - GPIO: con 40-pines.
  - Dimensiones: 65mm x 30mm x 5mm.
- Estructura de ensamble físico: Es la base del experimento y dependiendo de qué estructura se elija el sistema de movimiento. Se optó por una estructura impresa en 3D (Mecanismo Bob de código libre [4]), que servirá como caparazón o estructura externa de montaje para montar experimento. Ver Figura 6.
- Características de la cámara: Tipo: CSI, Resolución: 5 Mpx.
- Sistema de movimiento: Para trasladarse de un punto a otro contamos con un sistema de movimiento, basado en servomotores que en puntos específicos (articulaciones del mecanismo) ayudarán al movimiento de éste. Se proponen servomotores de 1.5 Kg. Véase la Figura 7.

#### Etapas desarrolladas en la computadora

Se realiza una normalización mediante un algoritmo para preparar los datos para la red neuronal que recibirá una biblioteca de imágenes de cuatro personas, cada



**Fig. 6.** Estructura física ensamblada del mecanismo Bob.



(a) La cámara usada.



(b) Servomotor de 1.5 Kg.

**Fig. 7.** Componentes adicionales.

persona de esta biblioteca tendrá cuatro fotos de  $320 \times 320$  píxeles con diferentes gestos faciales, estas se categorizarán mediante una biblioteca de OpenFace.

Con esta información se entrena una red neuronal que sea capaz de tener una alta fiabilidad para determinar si la persona está dentro de la base de datos y así dar una respuesta positiva o negativa. Esta red neuronal es del tipo supervisada y con activador tipo Relu, lo que hace que tenga un porcentaje de exactitud mayor [10].

### **Etapas desarrolladas en la Raspberry**

A partir de una foto tomada por el mecanismo cuando una persona se acerque a menos de un metro de distancia, empieza el procesamiento de la imagen primero normalizando la imagen e identificando el rostro por medio del algoritmo de la librería OpenCV.

Después de esto se procesará la imagen normalizada por una red neuronal ya entrenada en la computadora que procesará los puntos de interés entregados

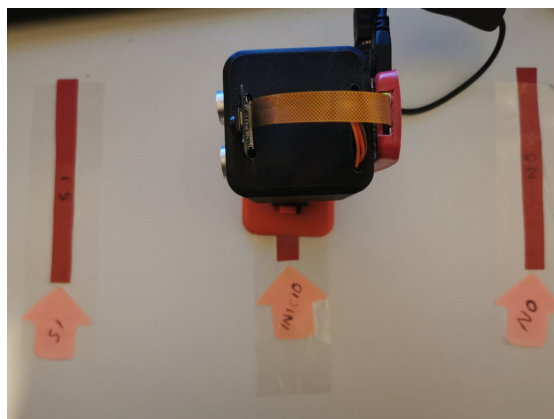
por el paso anterior y con eso poder obtener si esta imagen pertenece a alguna persona de la base de datos [10].

Como último paso, en el caso positivo de que el proceso detecte una persona que corresponda a la base de datos el mecanismo realiza movimientos para acercarse a la persona y en caso contrario realiza otro movimiento.

#### **4. Experimentación y resultados**

El experimento a realizar se centra en que el mecanismo tome la foto, reconozca el rostro y determine si éste pertenece a una persona que esté en la base de datos aprendida en el entrenamiento avance hacia adelante, caso contrario deberá retroceder. Todo este movimiento lo realizara mediante los servomotores instalados en la parte inferior del mecanismo.

Para ambos escenarios la posición inicial es la misma, la foto es tomada a una distancia menor a un metro y a una altura en que se pueda observar el rostro como se muestra en la Figura 8.



**Fig. 8.** Posición Inicial del mecanismo para el experimento.

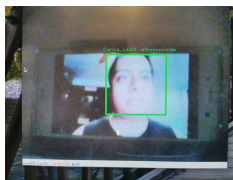
##### **Primer escenario**

En este escenario el mecanismo se encuentra con una foto de una persona a menos de un metro de distancia, en este punto toma la captura de la persona como se muestra en la Figura 9.

Cuando la cámara toma el rostro comienza el proceso de reconocimiento, esto ocurre en el algoritmo que ya ha sido entrenado previamente en la computadora y fue migrado a la Raspberry, en cuanto el algoritmo reconoce el rostro el mecanismo avanzara. Como se muestran en la Figura 10a y 10b.



Fig. 9. Captura del rostro para el primer escenario.



(a) Reconocimiento y Validación de una persona reconocida.



(b) Posición inicial.



(c) Posición Final.

Fig. 10. Ejemplo experimental con la fotografía capturada en un dispositivo móvil y el proceso de reconocimiento de rostro en el mecanismo.

De esta manera podemos ver como el algoritmo de reconocimiento pre-entrenado válida que es una persona que está en la base de datos y el mecanismo realiza correctamente su rutina después de esta validación como se puede ver en la Figura 10c.

### Segundo escenario

En este escenario el mecanismo se encuentra con una foto de otra persona que también está a menos de un metro de distancia y en este punto toma la captura de la persona como se muestra en la Figura 11.

Nuevamente la cámara captura el rostro y comienza el proceso de reconocimiento, en cuanto el algoritmo pre-entrenado reconoce que el rostro no coincide con alguno de la base de datos el mecanismo retrocederá. Como se muestran en la Figura 12a y 12b. De esta manera podemos ver como el algoritmo de



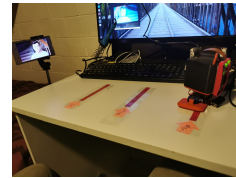
**Fig. 11.** Captura del rostro para el segundo escenario.



**(a)** Reconocimiento y Validación de una persona no reconocida.



**(b)** Posición inicial.



**(c)** Posición final.

**Fig. 12.** Proceso de reconocimiento del rostro en la pantalla y respuesta del mecanismo.

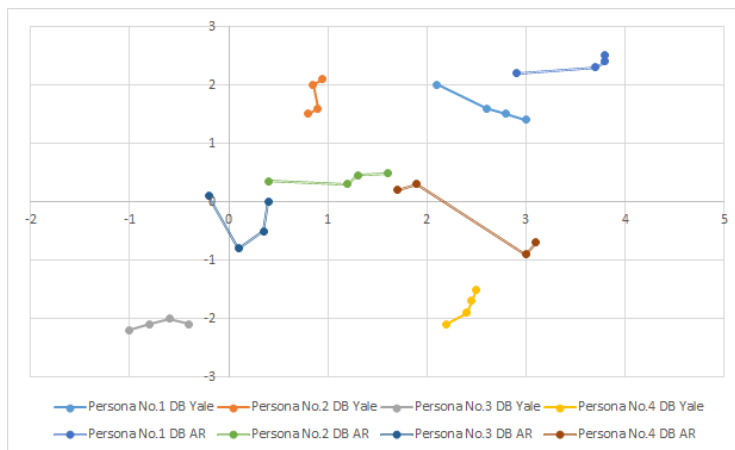
reconocimiento pre-entrenado válida que esta persona no está en la base de datos y el mecanismo realiza correctamente su rutina de retroceso después de esta validación como se puede ver en la Figura 12c.

#### 4.1. Resultados

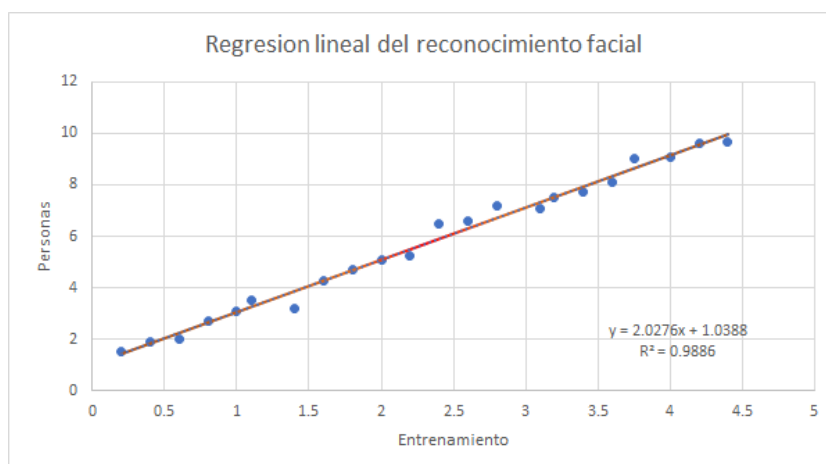
Presentamos los resultados obtenidos para el entrenamiento con las bases de datos consideradas, Yale y AR. Se puede apreciar en la figura 13 que los rostros se reconocen claramente y no se traslapan.

Se pudo validar ambos escenarios, el de reconocimiento de rostro cuando la persona es identificada como parte de la base de datos y en el que la persona es desconocida para el sistema, viendo que el mecanismo valido a la persona y después realizo la acción que tenía programada para cada escenario.

Un ejemplo del porcentaje de éxito del reconocimiento facial con el método seleccionado se puede observar en la siguiente Figura 14.



**Fig. 13.** Proyección de los vectores de los rostros, de la base de imágenes de Yale y AR.



**Fig. 14.** Regresión lineal

## 5. Conclusiones

En este trabajo hemos presentado una implementación de un sistema de reconocimiento de rostros basado en una red neuronal convolucional, pre-entrenada y de dominio público, en un sistema embebido portable, de código abierto y con librerías ampliamente conocidas por la comunidad científica de Visión por Computadora y Redes Neuronales artificiales. Todo el sistema se pudo montar en un mecanismo móvil abierto a la comunidad académica para su experimentación libre, con el propósito de poder interactuar con personas, infantes o jóvenes, para



su identificación por medio del rostro y la respuesta con movimiento asociado a consentimiento por parte de las personas.

Si bien consideramos que la implementación puede tener múltiples propósitos, hemos mostrado la factibilidad de la implementación de hardware y software concreto para una tarea práctica de reconocimiento de rostros, y de que el re-entrenamiento de nuevos rostros es posible con corto tiempo y toda la implementación a muy bajo costo en comparación con otros sistemas de mecanismos móviles del tipo en el mercado.

Como trabajo futuro estamos desarrollando la interfaz de nuestro desarrollo con un mecanismo móvil propietario, a fin de poder extender el uso de los mecanismos para nuevos propósitos y también extender su vida útil para fines educativos y empresariales. Así también estamos considerando la implementación de una metodología modificada que permita una actualización de la red entrenada con bases de datos masivas en un corto tiempo bajo el esquema cliente-servidor.

## Referencias

1. Ordieres-Meré, J., Limas, M., Ascacibar, F.J., Alba-Elías, F., González-Marcos, A., Pernía-Espinoza, A., Vergara, E.: Técnicas y algoritmos básicos de visión artificial. Recurso electrónico - En línea. Universidad de la Rioja (2006)
2. Fukunaga, K.: Introduction to statistical pattern recognition (2nd ed.). Academic Press Professional, Inc. (1990)
3. The Computer Vision Industry: <https://www.cs.ubc.ca/lowe/vision.html> (2020)
4. MakerBob Education: <https://www.thingiverse.com/thing:43708> (2019)
5. National Instruments Corp.: <http://www.ni.com/es-mx/shop/select/machine-vision-category> (2019)
6. Nalwa, V.S.: A Guided Tour of Computer Vision. Addison-Wesley Longman Publishing Co., Inc. (1993)
7. Pérez, M.A.C.: Reconocimiento robusto de rostros en ambientes dinámicos. Master's thesis, Universidad de Chile (2012)
8. Villanueva-Escudero, C., Villegas-Cortez, J., Zúñiga-López, A., Avilés-Cruz, C.: Monocular visual odometry based navigation for a differential mobile robot with android OS. Human-Inspired Computing and Its Applications, Lecture Notes in Computer Science, Vol. 8856, Springer International Publishing, pp. 281–292 (2014)
9. Olague, G.: Evolutionary Computer Vision: The First Footprints. Natural Computing Series. Springer-Verlag Berlin Heidelberg (2016)
10. Zúñiga Lopez, A., Villegas-Cortez, J., Avilés-Cruz, C., Rodríguez-Martínez, E., Ferreyra-Ramirez, A.: Reconocimiento de rostros por medio de Openface en una Raspberry PI. Research in Computing Science, 147, pp. 77–88 (2018)
11. Benavides-Alvarez, C.: Sistema no supervisado de clasificación de rostros con técnicas basadas en CBIR. tesis para obtener el grado de maestro en ciencias y tecnologías de la información. Universidad Autónoma Metropolitana (2015)
12. Benavides-Alvarez, C., Villegas-Cortez, J., Román-Alonso, G., Avilés-Cruz, C.: Face classification by local texture analysis through CBIR and surf points. Revista IEEE América Latina, 14 (2016)

13. Benavides Alvarez, C., Villegas Cortez, J., Román-Alonso, G., Aviles-Cruz, C.: Reconocimiento de rostros a partir de la propia imagen usando tecnica CBIR. In: X Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB), Merida, Extremadura, Spain, Universidad de Extremadura, pp. 733–740 (2015)
14. Sulis Setiowati, Z., Legya Franita, E., Ardiyanto, I.: A Review of Optimization Method in Face Recognition: Comparison Deep Learning and Non-Deep Learning Methods. In: ICITEE (2017)
15. Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E.: Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience* (2018)