



## Reconocimiento de Personas en Ambiente con Emisiones de Humo Usando Sensor Laser y Redes Neuronales Convolucionales desde Nube de Puntos 3D. Parte 1.

### Recognition of People in Environment with Smoke Emissions Using Laser Sensor and Convolutional Neural Networks from 3D Pointed Cloud. Part 1.

Abraham E. Gamarra Moreno<sup>1</sup>.; José A. Dávalos Pinto <sup>2</sup>

<sup>1</sup>. Universidad Nacional del Centro del Perú, Facultad de Ingeniería de Sistemas. Av. Mariscal Castilla N° 3909, El Tambo, Huancayo, Perú. e-mail: aegamarra@hotmail.com

<sup>2</sup> Pontificia Universidad Católica del Perú, Departamento de Ingeniería. Av. Universitaria 1801. Lima, Perú. e-mail: jdavalos@pucp.edu.pe

#### Resumen

Cuando ocurren incendios en Plantas industriales, los materiales siniestrados y el humo dificultan la identificación del personal. Los Drones con sensores laser y redes CNN posibilitan el reconocimiento de personas en tales ambientes. El objetivo del proyecto es el estudio de un sistema neuronal convolucional para el reconocimiento de personas en un ambiente con emisiones de humo (parte1) y para su implementación usa un sensor laser, una tarjeta controladora y un dron (parte2). El método empleado consideró la realización de un reconocimiento de personas usando una red CNN previamente entrenada con nube de puntos 3D. La prueba se realizó con Alexnet e imágenes de personas. Los resultados (parte1) muestran que una matriz de confusión del 97.5 % ha sido alcanzada. Basado en el estudio, se concluye que el sistema neuronal de reconocimiento de personas usando CNN en ambientes siniestrados presenta un comportamiento muy aceptable para su aplicabilidad.

*Palabras clave:* reconocimiento de personas, redes neuronales convolucionales, programación en entorno Matlab, sensor laser y dron.

#### Abstract

When fires occur in industrial plants, the damaged materials and smoke make it difficult to identify personnel. Drones with laser sensors and CNN networks enable the recognition of people in such environments. The objective of the project is the study of a convolutional neuronal system for the recognition of people in an environment with smoke emissions (part1) and for its implementation it uses a laser sensor, a controller card and a drone (part2). The method used considered the realization of a recognition of people using a CNN network previously trained with a 3D point cloud. The test was conducted with Alexnet and images of people. The results (part1) show that a confusion matrix of 97.5 % has been reached. Based on the study, it is concluded that the neuronal system of recognition of people using CNN in disaster environments presents a very acceptable behavior for its applicability

*Keywords:* recognition of people, convolutional neuronal networks, programming in Matlab environment, laser sensor and drone.

## 1. Introducción

Un área muy utilizada para el reconocimiento de objetos y personas es la visión artificial. La visión artificial se dedica a la identificación, localización y reconocimiento de objetos (Angulo, Romero, & Angulo, 2001). Una definición adicional menciona que la visión artificial por computador es la capacidad de la máquina para ver el mundo que le rodea, más precisamente para deducir la estructura y las propiedades del mundo tridimensional a partir de una o más imágenes bidimensionales (Pajares & De la Cruz, 2008).

El proceso de la visión artificial toma una escena tridimensional y obtiene imágenes monocromáticas o en color; luego las imágenes pueden segmentarse para obtener los bordes o regiones de los objetos; a continuación, se realiza el proceso de descripción donde se obtienen las propiedades en detalle de los objetos para luego utilizar este resultado en la aplicación de interés (Pajares & De la Cruz, 2008).

La obtención de las imágenes en la visión artificial se realiza a través de cámaras digitales y se sugiere que el ambiente este iluminado para mejores resultados, por lo que no sería adecuado en ambientes que emanan humo. “La iluminación de la imagen es un factor importante que suele afectar a la complejidad de los algoritmos de visión. La iluminación arbitraria del entorno no suele ser aceptable ya que se obtienen imágenes con bajo contraste, reflexiones especulares, sombras y detalles espúreos” (Fu, Gonzalez, & Lee, 1988).

La segmentación permite mejorar las imágenes para su posterior descripción. La segmentación se define como “el proceso que divide una imagen en objetos que sean de nuestro interés” (Fu, Gonzalez, & Lee, 1988). La mejora de las imágenes a través de la segmentación no tiene éxito en ambientes que emanan humo debido a que los objetos están ocultos detrás del humo.

Las técnicas de visión artificial para el reconocimiento de personas fallan cuando el ambiente emana humo, debido a que se tendría

una visibilidad baja. En (Expower, 2018) se menciona que “el humo también evita el paso de la luz dificultando el trabajo de los equipos de extinción y la extracción de los posibles heridos en el incendio”.

El uso de sensores LIDAR es una opción para la detección de personas u objetos en ambientes que no emanan humo y en trabajos de investigación previos se utilizaron estos sensores en combinación con una de las siguientes técnicas: algoritmos elementales de detección de personas (Hamme, Hebel, & Arens, 2016), algoritmos de detección de personas en una nube de puntos (Hammer, Hebel, & Arens, 2017) o la combinación de Modelo Mixto Gaussiano, clasificador bayesiano y el filtro de Kalman extendido (Tamás & Lazea, 2010).

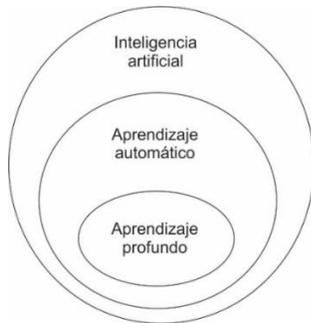
Debido a las limitaciones de las técnicas de visión artificial para el reconocimiento de personas, que se encuentran en ambientes que emanan humo y que tengan fuego; se han propuesto el uso de las siguientes técnicas : la tecnología de holografía digital en el rango infrarrojo para obtener imágenes a través del humo y las llamas (Locatelli, y otros, 2013), el uso del procesamiento térmico de imágenes estéreo que permite realizar el seguimiento de una persona y reconocimiento de gestos en ambientes llenos de humo (Kapusta, Beeson, & Patrick, 2016) y el uso de laser y sensores de sonido en combinación con un sistema de visión para el seguimiento de personas en ambientes lleno de humo (Sales, Marín, Cervera, Rodríguez, & Pérez, 2010).

Sin embargo, en este artículo el reconocimiento y ubicación de personas en ambientes que emanan humo, se realiza utilizando sensores LIDAR en combinación con el uso de redes neuronales convolucionales (CNN). Los resultados sirven para la búsqueda de personas en edificaciones incendiadas que todavía emanan humo.

## 2. Marco Teórico

## 2.1 Aprendizaje profundo

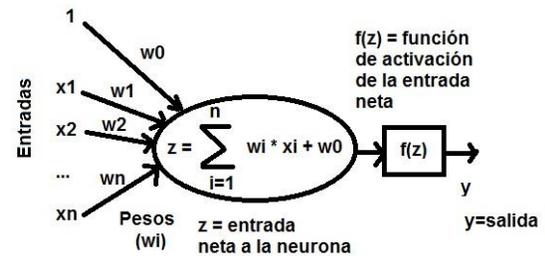
El aprendizaje profundo (Deep learning) es un área del aprendizaje automático, es una nueva forma de aprender a partir de los datos y pone énfasis en el aprendizaje desde capas sucesivas de representaciones (Chollet, 2018). La relación entre la inteligencia artificial, aprendizaje automático y aprendizaje profundo se muestra en la figura 1 (Chollet, 2018). El aprendizaje profundo se implementa utilizando una arquitectura de red neuronal y el término profundo se refiere a que la red puede tener hasta 150 capas (MathWorks, 2017).



**Figura 1.** Relación entre la inteligencia artificial, aprendizaje automático y aprendizaje profundo. Fuente: (Chollet, 2018).

## 2.2 Redes neuronales convolucionales

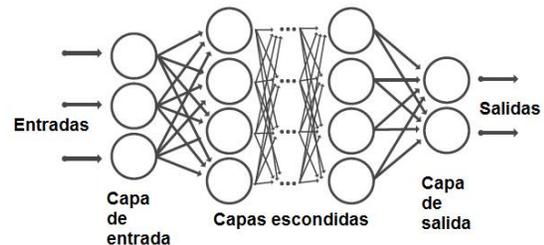
Las redes neuronales artificiales consisten de neuronas que tienen pesos entre sus conexiones y entradas activas denominadas “bias” (Joshi, 2017). Cada neurona recibe la entrada  $x_i$ , luego se calcula  $z$  que es igual a la sumatoria del producto de los pesos  $w_i$  por las entradas  $x_i$ , a continuación,  $z$  se envía a una función de activación  $f(z)$  para luego encontrar la salida  $y$  (figura 2).



**Figura 2.** Neurona artificial.

Fuente: elaboración propia

Las redes neuronales tienen una o más capas y la red neuronal multicapa tienen una capa de entrada, uno o más capas escondidas y una capa de salida (figura 3). Los algoritmos de aprendizaje de la red neuronal permiten ajustar los pesos de esta red.



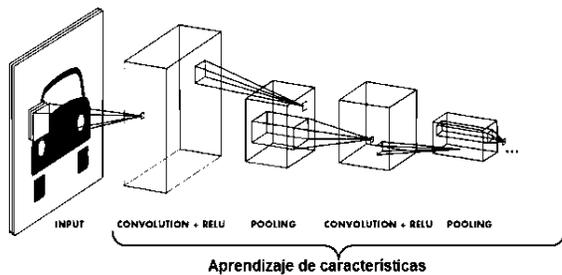
**Figura 3.** Red neuronal multicapa.

Fuente: (MathWorks, 2017).

Las redes neuronales convolucionales (CNN) es un tipo especializado de red neuronal para el procesamiento de datos que tiene una topología conocida, parecida a una matriz bidimensional. El nombre de red neuronal convolucional indica que la red emplea una operación matemática llamada convolución. La red neuronal convolucional está compuesta de una capa de entrada, muchas capas escondidas y una capa de salida. El objetivo de las redes neuronales convolucionales es tomar imágenes sin procesar en la capa de entrada para luego obtener una correcta clasificación en la capa de salida. La diferencia de las redes neuronales convolucionales con las redes neuronales ordinarias reside en el tratamiento de los datos y los tipos de capas que se utilizan (Joshi, 2017).

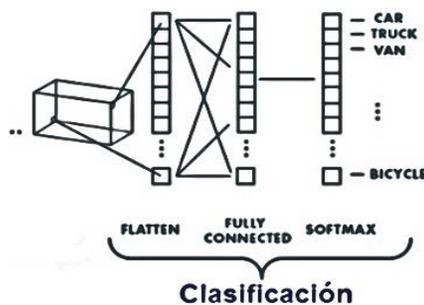
182

La red neuronal convolucional agrupa a las capas en detección de características y clasificación; en las capas de detección de características desarrollan uno de los tres tipos de operaciones: convolución (convolution), una capa de reducción (pooling), una capa de clasificación (rectified linear unit, ReLU), tal como se muestra en la figura 4. En la capa de clasificación se utilizan capas totalmente conectadas y la última capa utiliza la función softmax (figura 5) (MathWorks, 2017).



**Figura 4.** Capas input, conv, ReLU y pooling de CNN. Fuente: (MathWorks, 2017).

La capa convolucional calcula el producto punto entre los pesos y pequeñas áreas de la salida de capas previas, la capa de reducción o pooling reduce las dimensiones de las capas previas y la capa de clasificación o ReLU aplica una función de activación a la salida de la capa previa (Joshi, 2017).

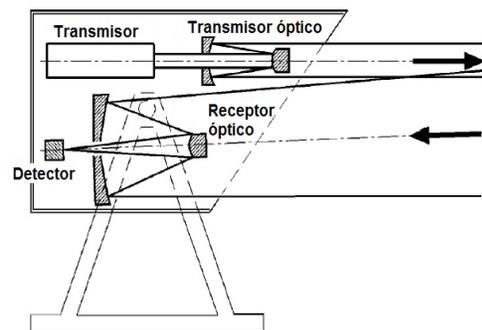


**Figura 5.** Capas fully connected y softmax de CNN. Fuente: (MathWorks, 2017).

La capa totalmente conectada (fully connected layer) calcula la clasificación que tiene cada clase en la salida.

### 2.3 Lidar

LIDAR, acrónimo (light detection and ranging), es una técnica de detección remota basada en laser (Campbell, 1997). Un sistema LIDAR consiste de: un transmisor laser; un transmisor óptico; un receptor óptico; un detector; y un sistema electrónico para adquisición de datos (Fujii & Fukuchi, 2005), como se muestra en la figura 6.

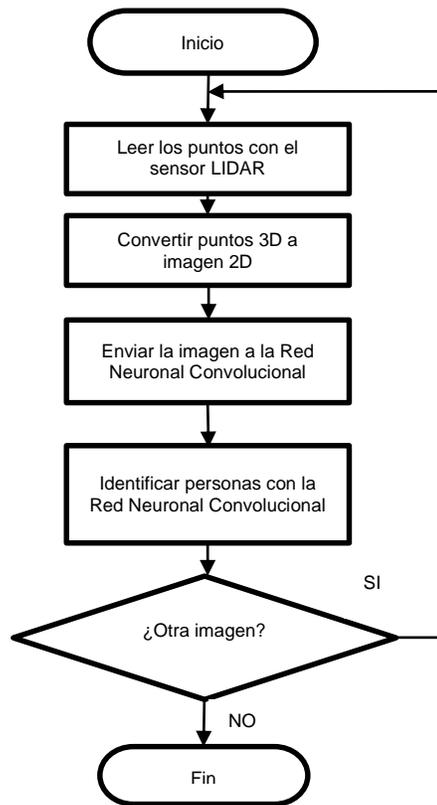


**Figura 6.** Sistema de sensado laser LIDAR. Fuente: (Fujii & Fukuchi, 2005).

El sensor LIDAR se utiliza para medir la distancia a la que se encuentra un objeto. El tiempo transcurrido  $t$  entre el pulso transmitido y el pulso reflejado desde el objeto que se encuentra a una distancia  $d$  y considerando que el pulso se traslada a la velocidad de la luz  $c$ , se tiene que la distancia es:  $d = \frac{1}{2} * c * t$ .

### 3. Metodología

El procedimiento seguido (parte 1 del estudio) para el reconocimiento de personas de la figura 7 utiliza sensores, tarjeta controladora y redes CNN.



**Figura 7.** Procedimiento para el reconocimiento de personas.

Fuente: Elaboración propia.

### 3.1 Entrenamiento de la red neuronal convolutiva

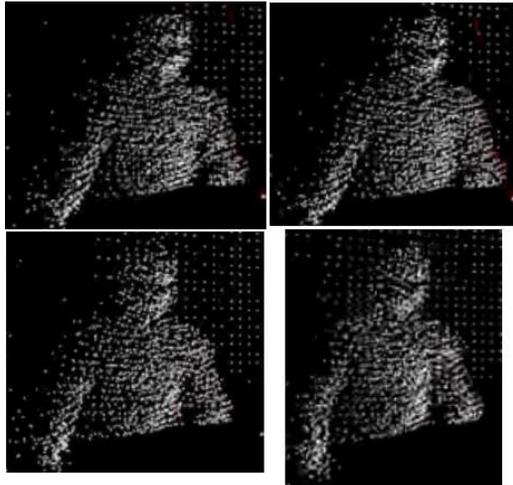
Para el entrenamiento de la red neuronal convolutiva no se ha partido desde cero, sino que se ha utilizado una transferencia de aprendizaje desde una CNN denominada Alexnet.

AlexNet es una red CNN preconfigurada que ha sido entrenada con aproximadamente 1,2 millones de imágenes de la base de datos ImageNet (MathWorks, 2018). AlexNet consta de 25 capas y puede clasificar imágenes en 1000 categorías de objetos (animales, teclado, mouse, taza de café, lápiz y muchos otros objetos). MATLAB tiene un programa específico toolbox que contiene la red entrenada y se puede transferir

el aprendizaje previo reentrenando la capa 23 (MathWorks, 2018). Las capas de AlexNet que se ha utilizado en la transferencia de aprendizaje son descritas a continuación:

- capa01: 'data', (image input [227x227x3]).
- capa02: 'conv1', (convolution 96[11x11x3] with stride [4 4] and padding [0 0 0]).
- capa03: 'relu1', (ReLU).
- capa04: 'norm1', (cross channel normalization with 5 channels per element).
- capa05: 'pool1', (max pooling 3x3 with stride [2 2] and padding [0 0 0]).
- capa06: 'conv2', (convolution 256[5x5x48] with stride [1 1] and padding [2 2 2]).
- capa07: 'relu2', (ReLU).
- capa08: 'norm2', (cross channel normalization with 5 channels per element).
- capa09: 'pool2', (max pooling 3x3 with stride [2 2] and padding [0 0 0]).
- capa10: 'conv3', (convolution 384[3x3x256] with stride [1 1] and padding [1 1 1]).
- capa11: 'relu3', (ReLU).
- capa12: 'conv4', (convolution 384[3x3x192] with stride [1 1] and padding [1 1 1]).
- capa13: 'relu4', (ReLU).
- capa14: 'conv5', (convolution 256[3x3x192] with stride [1 1] and padding [1 1 1]).
- capa15: 'relu5', (ReLU).
- capa16: 'pool5', (max pooling 3x3 with stride [2 2] and padding [0 0 0]).
- capa17: 'fc6', (fully connected 4096 layer)
- capa18: 'relu6', (ReLU).
- capa19: 'drop6', (dropout 50%).
- capa20: 'fc7', (fully connected 4096 layer).
- capa21: 'relu7', (ReLU).
- capa22: 'drop7', (dropout 50%).
- capa23: 'fc8', (fully connected 1000 layer).
- capa24: 'prob', (function softmax).
- capa25: 'output', (classification output with 'tench' and 999 other classes).

Para crear el prototipo de la identificación de personas en ambientes que emanan humo en Matlab, se ha realizado el reentrenamiento de la capa 23 de la red AlexNet utilizando 4 imágenes por cada categoría como se muestra en la figura 8.



**Figura 8.** Una categoría de imagen utilizada en el entrenamiento de la red CNN AlexNet.

Fuente: Elaboración propia.

### 3.2 Adquisición de datos (imágenes).

Para identificar a las personas que se encuentran en ambientes que emanan humo (parte 1 del estudio) se utiliza un sensor laser LIDAR-lite 3, una tarjeta controladora Arduino UNO y una Laptop con Matlab. El sensor obtiene una nube de puntos en 3D que representa a los objetos que naturalmente contienen personas que se encuentran en el ambiente.

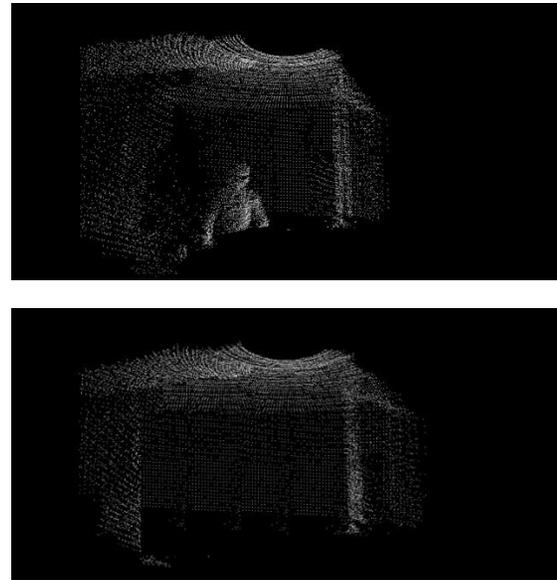
### 3.3 Conversión de imagen de puntos 3D a 2D.

La imagen que contiene a la nube de puntos en 3D se convierte a 2D y se realiza a través de un programa implementado en MATLAB.

### 3.4 Identificación de personas con red CNN.

Se ha implementado un prototipo en MATLAB que permite utilizar la red AlexNet reentrenada, el cual recibe las imágenes a clasificar. Para identificar si se encuentran personas dentro de las imágenes se envía a la red neuronal convolucional la imagen que incluyen personas e imágenes con

otros objetos en el ambiente como se muestra en la figura 9.



**Figura 9.** Imágenes enviadas a la red CNN que incluyen personas u otros objetos.

Fuente: Elaboración propia.

## 4. Resultados

Para obtener los resultados en la clasificación de personas en ambientes que emanan humo (parte 1 del estudio) se han enviado a la red neuronal convolucional 20 imágenes con ambientes que contienen personas (AP) y 20 imágenes con ambientes que no contienen personas (ASP).

La cantidad de imágenes enviadas a la red neuronal convolucional para cada clase se muestran en la tabla 1.

**Tabla 1.** Cantidad de imágenes utilizadas para la clasificación

| Características de la imagen     | Cantidad de imágenes |
|----------------------------------|----------------------|
| Ambiente con una persona         | 20                   |
| Ambiente con objeto (no persona) | 20                   |
| Total                            | 40                   |

Fuente: Elaboración propia.

185

Para evaluar el desempeño en la clasificación de personas que tiene la red neuronal convolucional en ambientes que emanan humo se ha utilizado una matriz de confusión.

Los resultados de la matriz de confusión en clasificación de personas se muestran en la tabla 2.

**Tabla 2.** Matriz de confusión en la clasificación

|                         |       | Predicción |          | Total | Exactitud del usuario |
|-------------------------|-------|------------|----------|-------|-----------------------|
|                         |       | AP         | ASP      |       |                       |
| Observación             | AP    | 19         | 1        | 20    | 95%                   |
|                         | ASP   | 0          | 20       | 20    | 100%                  |
|                         | Total | 19         | 21       | 40    |                       |
| Exactitud del productor |       | 100 %      | 95.238 % |       |                       |

AP(Ambiente con persona) y ASP(sin persona).

Fuente: Elaboración propia.

La discusión de los resultados de la tabla 2 es la siguiente:

- La diagonal principal muestra la cantidad de imágenes clasificadas en forma correcta, en este caso 19 para AP y 20 para ASP de un total de 40 imágenes.
- La cantidad de imágenes que no se clasificaron en forma correcta están fuera de la diagonal, en este caso esta en la intersección de la fila AP y de la columna ASP y existe 1 imagen clasificada en forma incorrecta.
- En la fila AP se observa que 19 imágenes de esta clase fueron clasificados por la CNN en forma correcta y 1 en forma incorrecta. El porcentaje de AP clasificados en forma correcta es de 95%.

- En la fila ASP se observa que 0 imágenes de esta clase fueron clasificados por la CNN en forma incorrecta y 20 en forma correcta. El porcentaje de ASP clasificados en forma correcta es de 100%.
- La exactitud global o el porcentaje de los datos que se han clasificado correctamente en la matriz de confusión es de 97.5%.
- El estadístico Kappa de la matriz de confusión es 0.95 y el efecto que tiene el azar en la clasificación de las imágenes.

## Conclusiones

De los resultados mostrados y de su discusión (parte 1 del estudio) se puede obtener las siguientes conclusiones:

- La red CNN clasifica los dos tipos de imagen con una exactitud global en la matriz de confusión del 97.5%, lo cual es aceptable para los objetivos del proyecto,
- La red neuronal convolucional AlexNet, permite transferir su aprendizaje a través del reentrenamiento de la capa 23 facilitando la clasificación de imágenes,
- Se ha usado de manera satisfactoria el estado del arte en redes neuronales convolucionales aplicadas a la clasificación de imágenes lo cual será de gran utilidad para abordar la parte 2 del estudio.

## Referencias Bibliográficas

- Campbell, M. (1997). *Sensor Systems for Environmental Monitoring*. Volume One: Sensor Technologies. Glasgow: Chapman & Hall.
- Chollet, F. (2018). *Deep Learning with Python*. New York: Manning Publications Co..
- Fujii, T. & Fukuchi, T. (2005). *Laser Remote Sensing*. Boca Raton: Taylor & Francis Group.



186

- Hammer, M., Hebel, M. & Arens, M. (2016). *Automated object detection and tracking with a flash LiDAR system*. s.l., SPIE.
- Hammer, M., Hebel, M. & Arens, M. (2017). *Person detection and tracking with a 360° lidar system*. Warsaw, SPIE.
- Kapusta, A., Beeson, P. (2016). Person Tracking and Gesture Recognition in Challenging Visibility Conditions Using 3D Thermal Sensing. New York, IEEE, pp. 1112-1119. DOI: 10.1109/ROMAN.2016.7745247
- Locatelli, M. et al. (2013). *Imaging live humans through smoke and flames using far-infrared digital holography*. Washington: Journal Optics Express.
- MathWorks (2017). *Introducing Deep Learning with MATLAB*. Natick: The MathWorks, Inc.
- Pajares, G. & De la Cruz, J. (2008). *Visión por Computador*. México.: Alfa y Omega.
- Sales, J. et al. (2010). Multi-Sensor Person Following in Low-Visibility Scenarios. Castellón, Sensors, pp. 10953-10966. DOI: 10.3390/s101210953.
- Tamás, L. & Lazea, G. (2010). Pattern recognition and tracking dynamic objects with LIDAR. Cluj-Napoca, ISR/ROBOTIK, pp. 278-283. Print ISBN: 978-3-8007-3273-9.

**A.E. Gamarra Moreno**, Doctor en Ingeniería (2012) de la Universidad Nacional Federico Villarreal – UNFV, Lima Perú; Magister en Informática (1997) de Pontificia Universidad Católica del Perú, Ingeniero Mecánico (2009) de la Universidad Nacional del Centro del Perú y graduado en Ingeniería de Sistemas (2017) en la Universidad Católica Los Ángeles de Chimbote, Perú. Profesor Asociado de la Universidad Nacional del Centro del Perú, Facultad de Ingeniería de Sistemas; miembro del Directorio de Recursos Humanos afines a la

ciencia, tecnología e innovación tecnológica (DINA-CONCYTEC, Perú). <https://orcid.org/0000-0002-3953-8014>.

**J.A. Dávalos Pinto**. Ingeniero colegiado, UNI-Perú, Magister en Ingeniería de Control y Automatización, PUCP-Perú y Doctor en Ingeniería de Sistemas, UNFV-Perú. Ha realizado estudios de perfeccionamiento en Control de Procesos Industriales, Microcomputación y Sistemas de Manufactura en SENAI-Brasil. Ha obtenido dos premios nacionales, es autor de un libro (formato digital) sobre Microcontroladores y posee una patente de invención. Realiza trabajos de Investigación en Modelamiento, simulación y control de Procesos Industriales los cuales son presentados en congresos y publicados en revistas indizadas. Es miembro del comité científico de la revista GTI de UIS-Colombia.

\*Para citar este artículo: Gamarra Moreno A.E.; José A. Dávalos Pinto J.A. Recognition of People in Environment with Smoke Emissions Using Laser Sensor and Convolutional Neural Networks from 3D Pointed Cloud. Part 1. Revista Bistua. 2019 17(1):179-186

+ Autor para el envío de correspondencia y la solicitud de las separatas: Gamarra Moreno A.E.. Universidad Nacional del Centro del Perú, Facultad de Ingeniería de Sistemas. e-mail: aegamarra@hotmail.com

Recibido: Marzo 21 de 2018

Aceptado: Agosto 17 de 2018