

Monografía
Reconocimiento De Objetos Usando Técnicas De Inteligencia Artificial

Wilmer Giraldo Loiza

Facultad de Ingenierías- Ingeniería de Sistemas Y Computación.

Universidad Tecnológica De Pereira

Pereira 2020

Monografía

Reconocimiento de Objetos Usando técnica de Inteligencia artificial

Wilmer Giraldo Loaiza

Trabajo de grado presentado como requisito para optar al título de
Ingeniero de sistemas y computación.

Director:

Carlos Augusto Meneses Escobar,
Ingeniero De Sistemas y computación

Universidad Tecnológica De Pereira

Pereira2020

Contenido	
Índice de Tablas	6
Dedicatoria	7
Agradecimientos.....	8
Introducción	9
Capítulo 1	10
Generalidades	10
Planteamiento del Problema	10
Objetivos	11
Justificación	12
Capítulo 2	13
Estado del arte.....	13
2.1 Marco Referencial.....	13
2.2 Marco Conceptual.....	29
2.3 Marco Teórico.....	30
2.3.1 Visión Artificial	30
Visión Artificial.....	¡Error! Marcador no definido.
2.3.2 Redes Neuronales (CNN: neural network)	42
Red Neuronal Convolutacional	43
2.3.3 Máquinas núcleo, máquinas de vectores soporte (SVMs).....	45
a- Aprendizaje Bayesiano.....	46
b- Definición de red bayesiana:.....	46
2.3.5 Árboles de decisión	47

a- Bosque aleatorio.	47
b- AdaBoost.....	47
2.3.6 Técnica de visión por computadora.....	47
Fortalezas de técnicas.	47
Reconocimiento de objetos	48
Glosario	51
Conclusiones.....	50

Tabla de Ilustraciones

Ilustración 1.....	13
Ilustración 2.....	15
Ilustración 3.....	17
Ilustración 4.....	18
Ilustración 5.....	19
Ilustración 6.....	20
Ilustración 7.....	22
Ilustración 8.....	23
Ilustración 9.....	24
Ilustración 10.....	26
Ilustración 11.....	28
Ilustración 12.....	31
Ilustración 13.....	31
Ilustración 14.....	32
Ilustración 15.....	33
Ilustración 16.....	33
Ilustración 17.....	34
Ilustración 18.....	34
Ilustración 19.....	35
Ilustración 20.....	35
Ilustración 21.....	37
Ilustración 22.....	37
Ilustración 23.....	38
Ilustración 24.....	39
Ilustración 25.....	41
Ilustración 26.....	43
Ilustración 27.....	44
Ilustración 28.....	46

Ilustración 29..... 49

Índice de Tablas

Tabla 1..... 45

Tabla 2..... 48

Dedicatoria

Dedico este trabajo a mis Padres porque siempre me dan su apoyo.
A mis hijas y sobrinos porque son mi motivación.
A mi hermana, esa mujer guerrera que siempre me acompaña.
A mi esposa, mujer que adoro y que es mi compañera de vida.

Agradecimientos

Agradezco primero a Dios por ser mi guía en mi vida y que me llena de bendiciones y muchas alegrías.

A mis familiares y amigos que de una u otra manera me alentaron a no darme por vencido a alcanzar esta meta tan importante y que fue muy difícil de alcanzar, quiero ser un ejemplo para aquellos que creen que, si se puede desde que se tenga voluntad, ganas de salir adelante y lo más importante personas que te den el aliciente cuando uno sienta decaer.

Al ingeniero Carlos Augusto Meneses Escobar por darme su apoyo como director y guía en mi proyecto, muchas gracias.

Gracia a todos.

Introducción

Con el avance tecnológico de las últimas décadas, se han desarrollado múltiples técnicas para la clasificación y reconocimiento de imágenes, las cuales han servido de base para la solución de diversas necesidades de identificación de imágenes que se usan en la actualidad en aplicaciones comerciales y de seguridad.

El desarrollo soluciones para el reconocimiento de imágenes y clasificación de objetos emulan la información visual que recibe un ser humano, donde las cámaras ópticas hacen el papel del ojo humano, este procesa la información y por consiguiente la divide en una tarea simple capturar el objeto y reconocerle.

El reconocimiento de objetos en el campo de la visión por computadora de la inteligencia artificial es actualmente utilizado en el ambiente industrial, sistemas de conducción autónoma en vehículos, seguridad, Smartphone, consolas de videojuegos, seguridad y vigilancia, aplicaciones móviles o nuevos dispositivos móviles inteligentes.

Estas técnicas se usan en el presente como ayuda para automatizar procesos y servir de solución en control asistido, por ejemplo para conducir vehículos de forma automatizada y con total autonomía se puede evidenciar en marcas de automóvil como: F015 Luxury in Motion. [1, p. 1]

Estos grandes desarrollos han abierto las puertas a avances en el mundo de la seguridad, un ejemplo de esto son los vehículos aéreos no tripulados (VANT). Estos drones están equipados con múltiples dispositivos, entre ellos y de gran importancia las cámaras digitales, este es un ejemplo de que, entre más inteligencia artificial, mayor autonomía.

EL presente documento contiene la descripción de técnicas que se usan en el análisis de imágenes digitales para el reconocimiento de elementos de interés que se encuentran contenidos en ellas.

Capítulo 1

Generalidades

1.1. Planteamiento del Problema

El sentido de la vista del ser humano es uno de los sistemas más complejos dentro del mismo organismo, este percibe color, distancia, forma y dimensión, estos distinguen desde 2 hasta 10 millones de colores, con una pupila de se puede contraer o reducir haciendo que el enfoque pueda cambiar en milisegundos, la panorámica de la visión puede ser hasta de 180° lo que pasa esa información a la corteza visual y por consiguiente a un procesamiento e interpretación. [1]

La emulación que hace el proceso de reconocimiento de imágenes mediante los software que constantemente ha evolucionado con el fin de construir una visión artificial, utilizando condiciones controladas aún tiene como meta analizar las diferentes variaciones que se puedan presentar en la expresión facial, enfoque, iluminación del medio y todo objeto que es captado a través de un sistema de visión artificial, siendo este fundamental para un reconocimiento correcto y eficaz.

Los autores Russell y Norvig plantean el problema que se expone hoy: “El problema del reconocimiento visual de objetos es generalmente sencillo para las personas, pero se ha probado que es muy complejo para los computadores. Uno quiere ser capaz de identificar la cara de una persona a pesar de las variaciones de iluminación, postura respecto a una cámara y expresiones faciales.

Cualquiera de estos cambios causa diferencias generales en los valores de brillo de un píxel, así que realizar sencillamente una comparación de píxeles es poco probable que funcione. Cuando se desea reconocer ejemplos de una categoría, como por ejemplo un «coche», se debe hacer frente a las variaciones dentro de la categoría. Incluso el restringido problema del reconocimiento de los números escritos a mano en los códigos postales resulta ser un desafío.”. [2]

En la actualidad hay muchos métodos que son utilizados para la identificación de usuarios, contraseñas, número de identificación personal, pasaportes, métodos que se pueden perder en cualquier momento. Con este documento se pretende tener la importancia de actualizar conocimientos y compilar en esta monografía documentos actualizados que sirvan de consulta para conocer acerca del desarrollo del reconocimiento de objetos del campo de la visión por computadora en la inteligencia artificial.

1.2 Objetivos

1.2.1. Objetivo General.

Construir un documento monográfico, que reúna técnicas de inteligencia artificial para realizar el reconocimiento de objetos a partir de imágenes digitales.

1.2.2. Objetivos específicos.

- Hacer una revisión del estado del arte en las temáticas que se enmarcan en el reconocimiento de objetos en la inteligencia artificial.
- Hacer una clasificación de las técnicas de inteligencia artificial, utilizadas para el reconocimiento de objetos en imágenes digitales.
- Construir un documento monográfico del estado del arte del reconocimiento de objetos usando técnicas de inteligencia artificial.

1.3 Justificación

La clasificación y reconocimiento de objetos es utilizada en la actualidad por empresas con etapas productivas, empresas que fabrican automóviles, tienen robots que realizan tareas reconociendo objetos o realizando tareas específicas y sistematizadas. Esta monografía se basa en explicar el desarrollo de reconocimiento de objetos, las técnicas y algoritmos que se encargan del proceso desde la captura hasta la interpretación de la imagen.

Hace énfasis en las principales técnicas de procesamiento y reconocimiento de imágenes utilizados en las investigaciones plasmadas en el texto, **Sift** (Scale-invariant feature transform) extrae características de la imagen, para reconocerla en una base de datos, **SURF** (Speeded Up Robust Features) usado como algoritmo de extracción de puntos de interés en el reconocimiento de imágenes, **Deep Relation Graph Learning** : caracterización de imágenes aéreas mediante visión infrarroja, **BSIF** (binarized statistical image features) para reconocimiento de iris, algoritmos de **aprendizaje RF de 18 árboles** para vehículos aéreos no tripulados, **clasificador K** empleado para clasificación de datos en investigación de análisis facial térmico para la detección de engaños; **aprendizaje automático** con algoritmos para la detección de tamaño y color de imágenes, **HOG** (Histograms of Oriented Gradient) empleado para detección de animales y prevención de colisiones, **redes neuronales convolucionales 3D** con el fin de obtener información en estructura tridimensional en búsqueda de campos donde un helicóptero pueda aterrizar; **Depth of field (DOF)** o profundidad de campo como configuración óptica que referencia al rango espacial y de distancia con fotografía macro y micro entre otros.

Los softwares programados para realizar tareas concisas dentro de una compañía, usan sensores que reciben una imagen del objeto el cual utiliza técnicas de inteligencia artificial, se asume la evolución de las tecnologías en el presente año llevada a las empresas para trascender las tareas que realizan de forma Manual. [3]

Capítulo 2

Estado del arte

2.1 Marco Referencial

- a) *Reconocimiento de objetos ISAR del mundo real y descubrimiento de relaciones mediante el aprendizaje de gráficos de relaciones profundas.*

La investigación exhaustiva de estos autores sobre el reconocimiento de objetos de radar de apertura sintética inversa (ISAR) conocido por clasificar objetos multimodales como helicópteros, aviones, embarcaciones, automóviles, etc., tiene un problema de baja resolución a comparación con las imágenes infrarrojas y naturales, lo que presenta dificultad al reconocer imágenes que cambian de forma respecto a los otros vehículos.

Se evidencia un prometedor avance en las redes neuronales convolucionales profundas (DCNN) que tiene el inconveniente de ser limitado al modelar transformaciones que pueden pasar en las estructuras geométricas fijas de los módulos CNN. A partir de esta problemática el artículo propone un método que supera a los métodos de última generación de reconocimiento de objetos ISAR, basado en Deep Relation Graph Learning con el cual se producen conjuntos de datos de imágenes y las características del área y caracterización semántica intermodal con atributos para aprendizaje de asociación de los mismos. [12]

La siguiente imagen muestra como se identifican areas de interés donde se puede evidenciar en la figura 1 objetos dentro de las imágenes ISAR son más complejos que los que están dentro de las imágenes naturales e infrarrojas. Siendo ISAR el numero 2 y 4 de izquierda a derecha. La imagen 1 y la 3 pertenecen a visio infraroja, que apoya con mejor visualizacion el objeto

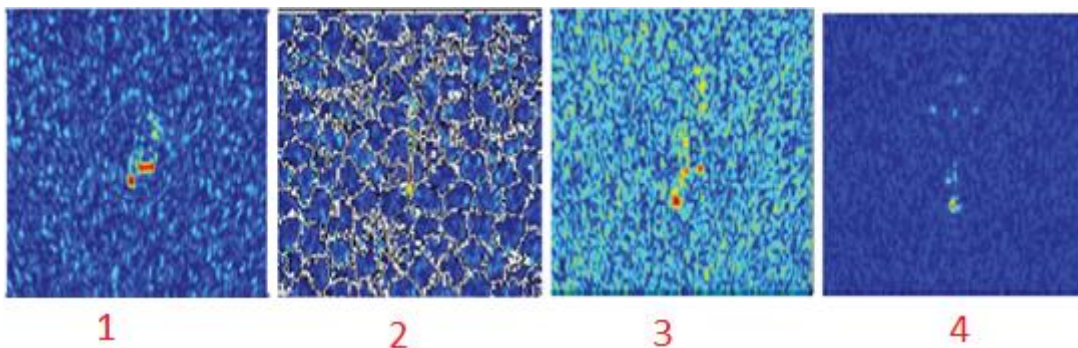


Ilustración 1

Autor: [12a]

b) Reconocimiento de objetos basado en RGB-D mediante redes neuronales convolucionales multimodales: una encuesta

La investigación se basó en los fundamentos claves para la rama de la robótica, centrándose en sensores a bajo costo que daban como resultado imágenes en RGB de alta calidad, este reconocimiento de objetos se basa en RGB-D de última generación basado en el aprendizaje profundo de las redes neuronales convolucionales (CNN), la investigación muestra dos cuestiones clave, una es la deficiencia de datos de entrenamiento y la fusión multimodal, identificando los métodos de vanguardia de reconocimiento de objetos que se basan en RGB-D y nuevas aplicaciones de aprendizaje multimodal ya sea de reconocimiento de voz, objetos.

En el reconocimiento de objetos existen varias limitantes, entre esas es el mundo 3D se proyecta en 2D y por lo tanto se pierden datos, como apariencia de objetos, iluminación, etc. Las empresas que en los últimos años han utilizado los sensores RGB-D más conocidas son: Kinect de Microsoft, Xtion PRO, Realsense de Intel, exponiendo sus imágenes de alta calidad respecto a la profundidad. Se ha vuelto más fuerte la caracterización de imágenes gracias a empresas como ImagiNet por sus sistemas informáticos de alto rendimiento como unidades de procesamiento de gráficos (GPU) que encontraron solución a problemas de reconocimiento de objetos en RGB-D basándose en MMCNN. (redes neuronales convolucionales multimodal) entrenando las deficiencias concretamente en función multimodal, clasificando métodos con esquemas semi-supervisados frente a los datos entrenados y no entrenados. [13]

c) SLAM semántico basado en la detección de objetos y Octomap mejorado.

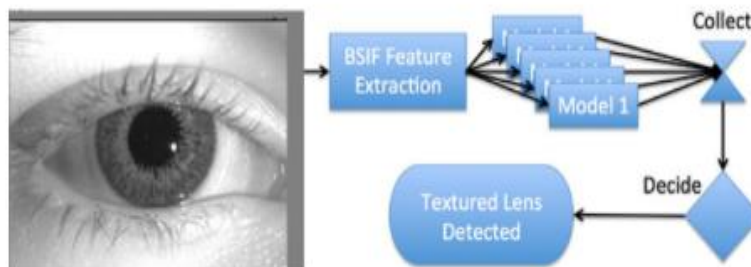
En este artículo se propone el sistema semántico SLAM (Simultaneous Localization and Mapping) para detección de objetos y reconocimiento de estos, de la mano de un módulo de ubicación exacta de la cámara y su respectivo mapeo en tiempo real. El módulo de detección de objetos y localización se completan con el fin de crear Octomap basado en algoritmos de rasterización de línea rápida. Lo que termina por abordar tres campos diferentes: el mapeo, ubicación y reconocimiento de objetos. Estos robots con SLAM tradicional pueden dar un mapa geométrico, pero no pueden dar un significado semántico como ejemplo: la taza es verde, pero el SLAM semántico RGB-D puede contener información de los obstáculos, su ubicación puesto que ofrece mapas en 3D y detección de objetos mediante el método de aprendizaje profundo.

Se concluye que el SLAM semántico es más preciso en los tres aspectos mencionados frente al ORB-SLAM siendo el optimizador semántico más rápido y preciso en la elaboración de mapas Octomap, reafirmando la eficiencia en la detección de objetos usando YOLO (You Only Look Once) que detecta objetos y generando fotografías, mapas semánticos y así filtrar las características del producto en cuestión, sumando información semántica. [14]

d) *Detección sólida de lentes de contacto con textura en el reconocimiento de iris mediante BSIF (binarized statistical image features).*

Esta investigación tuvo un enfoque en la creación de un algoritmo que detectara mediante lentes de contacto cosméticos con reconocimiento de iris. Pero hasta ahora no hay sistema de reconocimiento de objetos que capture la textura natural de un iris o de detección de lentes texturizados en el ojo.

El estado del arte de esta investigación expone la detección de lentes de contacto refiere tres problemas, el primero es que no se requiere una segmentación del iris para lograr la detección de lentes de contacto texturizados, el segundo problema que se presenta al querer hacer un algoritmo robusto es que los factores específicos del sensor no son entrenados con la precisión como si fuese un solo sensor. El entrenamiento de un sensor de sistema biométrico debe estar sujeto a modelos adicionales con el fin de mantener una superioridad referente a la detección de lentes texturizados; y la tercera problemática presentada es que el CCR del detector de lentes cosméticos al introducir un nuevo lente que no había detectado se cae con gran facilidad. [15]



La cámara adquiere una imagen del iris. La extracción de características BSIF se aplica a la región del iris. Un conjunto de modelos vota a favor o en contra de que sea una lente cosmética. El voto mayoritario se utiliza como clasificación final. En este caso, el conjunto lo clasifica como una lente cosmética / texturizada. [Ver más](#)

Ilustración 2

Autor: [15]

e) Segmentación de imágenes semánticas basada en ontología mediante modelos de mezcla y varios CRF (Conditional Random Field)

La estrecha relación que hay entre la segmentación semántica y la segmentación y clasificación de imágenes lleva a este estudio a dividir las imágenes que se les asignan etiquetas semánticas predefinidas, que usan características demostradas por píxeles o superpíxeles alimentados como campo aleatorio condicional (CRF) demuestra una discriminación poco coherente con el objeto y degrada el desempeño etiquetado; por este motivo se propone el enfoque de segmentación semántica de imágenes (OBSIS) que se basa en ontología, modelando la segmentación y detección de los objetos; este modelo acoge el espacio visual de bajo nivel a un nivel semántico intermedio, sus características se aprenden de forma individual y aprende.

El espacio visual de dimensionalidad reducida se convierte en espacio semántico combinando los procesos de Dirilecht y CFR, esto incluye las etiquetas intermedias para etiquetado de las regiones, así que se usan varios valores en la característica para darle una variabilidad semántica a la misma. La ponderación del método propuesto causa diferencias entre los objetos o regiones, también tiene en cuenta las similitudes viendo desde el punto de vista del aprendizaje humano. [16]

f) Detección de baches y grietas en el pavimento de asfalto basada en imágenes multiespectrales de vehículos aéreos no tripulados

Durante el uso del asfalto en las carreteras se puede deteriorar la capa de pavimento. en este documento se expone formas de teledetección de datos de averías en el pavimento desde vehículos aéreos no tripulados (UAV) que captan imágenes multiespectrales con características espaciales y espectro del objeto detección de rango y radar. Mediante algoritmos de aprendizaje automático, red neuronal artificial y bosque aleatorio, con estos elementos se consigue captar los elementos creando una imagen binaria que nos muestra algoritmos morfológicos, la medida de abertura, dilatación del terreno o erosión, área, longitud, intensidad.

El enfoque de este estudio se desarrolló en la ciudad de shihezi donde se caracterizaron las grietas y baches con los algoritmos de aprendizaje RF de 18 árboles y MSI (imágenes multiespectrales) y RGB, se muestra que hay una escala mínima de objetos de deterioro del asfalto, cuando las grietas son más pequeñas que las permitidas en la escala pueden pasar por alto. [17].

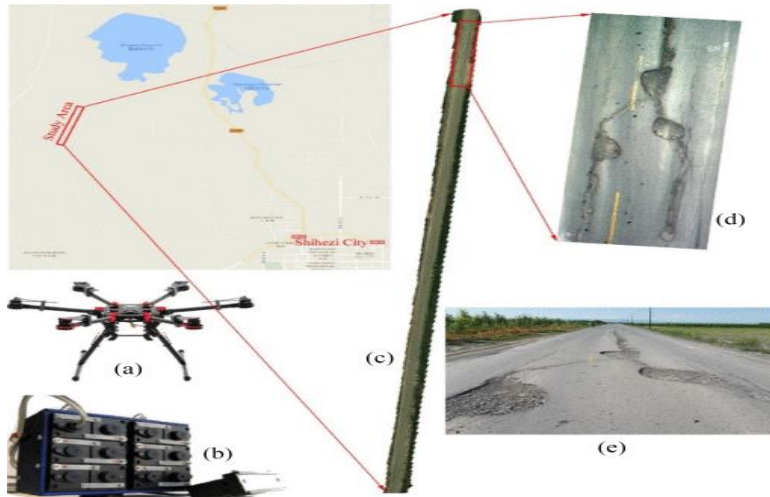


Ilustración 3

Autor: [17a] En esta imagen se muestra que se necesitan tres elementos para que el algoritmo detecte las grietas y la extensión de esta, primero el vehículo aéreo no tripulado con visión como el cuadro 1 a la izquierda, conexión continua a servidor web y por ultimo baches o grietas que pueda evidenciar el algoritmo.

g) Un método de aprendizaje profundo en conjunto para la clasificación de tipos de vehículos en sensores de vigilancia visual del tráfico

El aumento de los automotores a nivel mundial requiere que los sistemas de vigilancia, sean más rápidos y con mayor capacidad de almacenamiento, acompañado de clasificación eficaz y semántica ej.: objeto detenido o en movimiento y la velocidad en que este se mueve, Como objetivo se tiene dar etiqueta a cada imagen sobre que categoría determinada es dentro de los datos.

los algoritmos propuestos de clasificación de imágenes maximizan el porcentaje de predicción correcta de detección de automóviles, camiones o autobuses, desde sensores de vigilancia desde diferentes ángulos. Se propone integrar las redes neuronales profundas en dos etapas, la primera aumentar los datos de muestreo para equilibrar el desequilibrio de equilibrio de datos; la segunda etapa las redes convolucionales hacen la estructura de arquitectura de los datos enseñados y aumentar el entrenamiento con aportes de los datos adquiridos por medio de cámaras de trafico de MIOvision. [18]

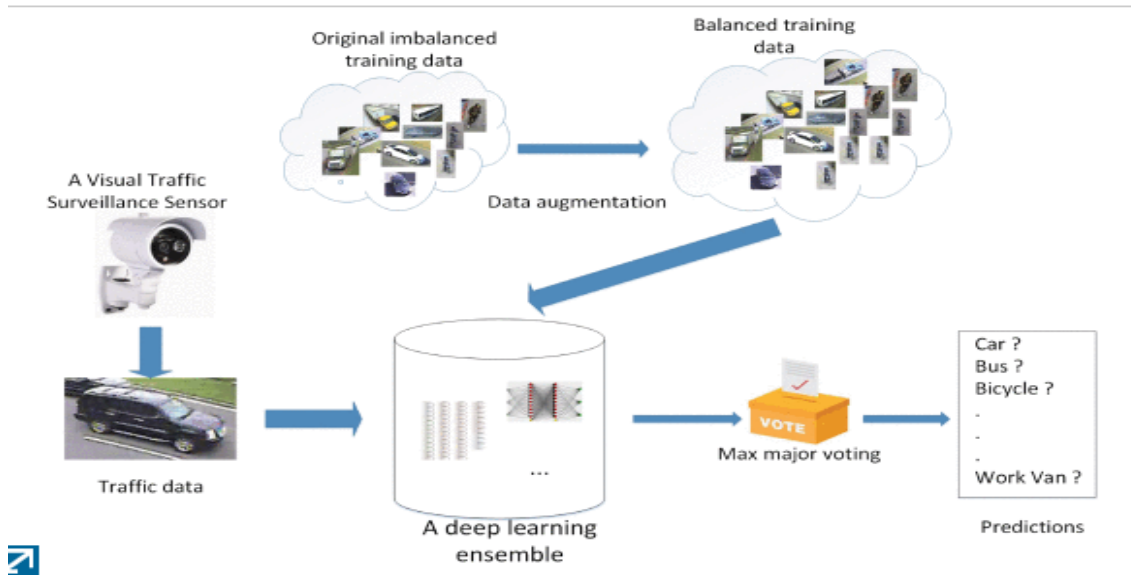


Ilustración 4

Autor: [18]

- i) *Un sistema de seguimiento de objetos en tiempo real basado en bordes direccionales que emplea generación de ubicaciones de candidatos múltiples.*

Este sistema de seguimiento se basa en un algoritmo de aprendizaje en línea para un rendimiento de seguimiento sólido, una matriz de puertas programables en campo (FPGA) esta procesa video en resolución 640 x480 pixeles 0. 1 ms en modo de alta velocidad y 0,8 ms en modo de alta precisión con frecuencia de trabajo de 60 MHz y suministra la ubicación del objeto en tiempo real de esta forma el aprendizaje en línea mejora el rendimiento de la matriz, todo esto para extraer las características del objetivo Y la transmisión de los datos.

Este algoritmo construido en la FPGA mejoro el rendimiento de la representación de los objetos y su ubicación, pero se discutió el alto costo que tiene la implementación del mismo para que funcione en tiempo real, sin embargo, se hace seguimiento al hardware en la matriz de puertas amigables en campo (FPGA) y se integra con tecnología de integración a muy gran escala (VLSI) lo que logra un mayor rendimiento y baja los costos computacionales. [19]

j) *Análisis facial térmico para la detección de engaños*

El análisis facial térmico propuesto por este estudio es la misma tecnología que se utiliza para detectar el aumento de calor en la cara de una persona en la zona periorbitaria, el detector de mentiras que normalmente se conoce como polígrafo es invasivo, pero el sensor térmico de esta investigación está programado para precisar en el cambio del comportamiento de la fisiología humana.

La culpa genera cambios fisiológicos y comportamentales por el cambio de flujo sanguíneo, frecuencia respiratoria y cardíaca; con esta metodología se detecta automáticamente los engaños. Este fue validado mediante 492 respuestas, 243 de ellas verdades y 249 mentiras donde se clasificaron las respuestas térmicas, dando como resultado una predicción del 87 % de las verdades o mentira, se utilizó el clasificador K para clasificación de las estrategias de datos. [20]



Configuración experimental: las grabaciones térmicas del sujeto se capturan mientras el examinador hace las preguntas.

Ilustración 5

Autor: [20]

k) Sistema de reconocimiento facial basado en imágenes con color

El algoritmo de aprendizaje AdaBoost desarrollado en esta investigación, está enfocado en el reconocimiento facial en 2D con orientación facial, la primera fase se detecta el rostro humano sobre el que se usa el AdaBoost, la segunda fase de la investigación se implementó un clasificador para reconocimiento del individuo usando la distancia euclidiana y de mahalanobis.

Los resultados expuestos reafirman la importancia que tiene el color en el sistema, con la base de datos RGB y escala de grises, el estudio tiene un promedio de 40,02% de que los colores usados que mejoraron el reconocimiento facial con planos HSV, pero mejora la calidad al reconocer el rostro con la técnica de plano RGB como procesamiento digital de imágenes como se puede evidenciar en la siguiente imagen: [21]

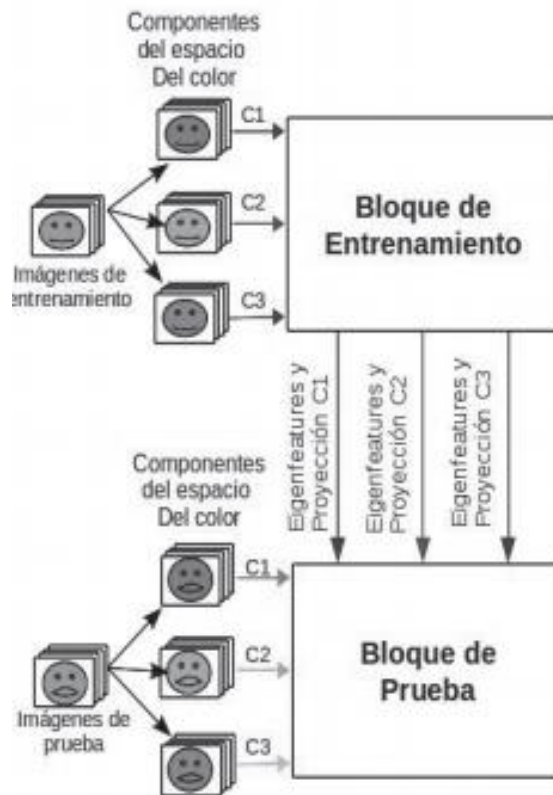


Ilustración 6

Autores: [21]

l) Algoritmos para la detección de tamaño y color de imágenes de teléfonos inteligentes de heridas crónicas para aplicaciones sanitarias

La aplicación móvil SmartWoundCare desarrollada para teléfonos móviles IOS y Android hace registro electrónico mediante imagen de la herida del paciente y hace seguimiento de la progresión gráfica, determinando el tamaño y el color de la herida mediante la cámara del teléfono o la Tablet; los sensores internos de la cámara miden con precisión mediante tres componentes: primero se obtienen las dimensiones del objeto en la imagen, el segundo componente es calibración de la cámara que convierte la imagen 2D a una Imagen 3D y un tercer algoritmo que extrae el rango de colores de formato rojo, verde y azul y convierte a colores amarillo rojo y negro.

Mediante el aprendizaje automático y sistemas expertos, con el fin de mejorar la correlación de la herida con la etapa en que se encuentra, así las escaras crónicas o úlceras de presión puede diagnosticar el crecimiento progresivo. [22]

m) Un sistema práctico de detección de animales y prevención de colisiones que utiliza una técnica de visión por computadora

Este artículo expone un método de bajo costo para la detección de animales en las carreteras, con el fin de disminuir el índice de accidentes de tránsito a través de la técnica de visión por computadora con una base de 2200 imágenes positivas y negativas, siendo probado con varios videos de vehículos movimiento, este método alerta a el conductor que va a 35 km/h en 2 segundos tras haber sido detectado el animal, aunque detecta el animal cuando el vehículo va a mayor velocidad, el conductor no tiene el tiempo y espacio para frenar.

Este algoritmo puede detectar animales en HOG (Histograms of Oriented Gradient) y clasificador en cascada detectando al animal en diferentes condiciones alcanzando una precisión de 82,5% a la distancia estimada anteriormente, el método propuesto se ampliaría fácilmente no solo en carreteras de la india, también de partes del mundo donde se pueden presentar inconvenientes con diferentes animales, obviamente cambiando las imágenes de un animal a otro dependiendo de los requerimientos del país. [23]

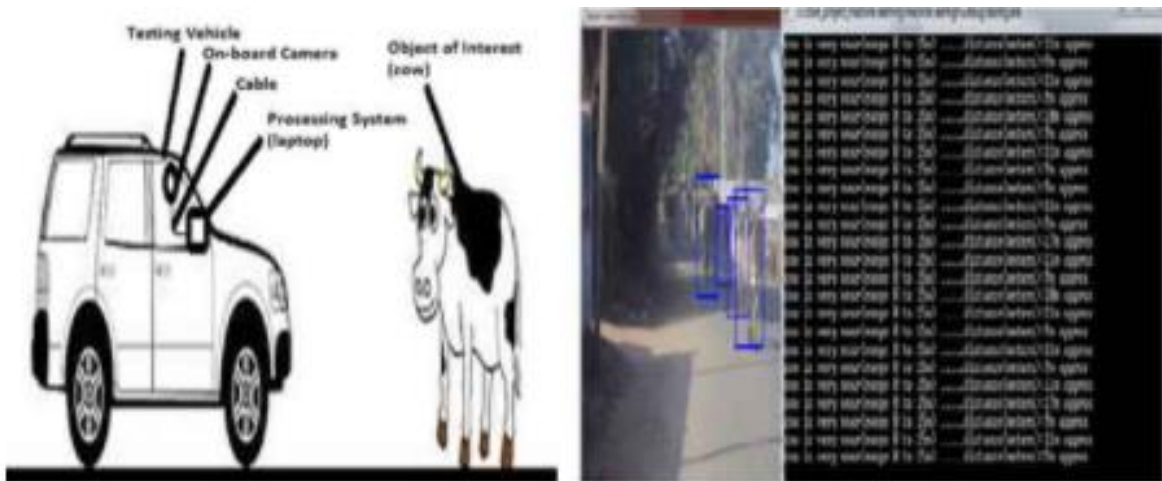


Ilustración 7

Autor:[23a]

En esta imagen se demuestra como la cámara del carro detecta la vaca a 20 km\h y pasa la información al sensor de luz que remite el sonido o la luz al conductor.

n) Salient Object Detection and Segmentation via Ultra-Contrast

Este estudio de objetos sobresalientes tiene la teoría de que los algoritmos tradicionales de objetos salientes se centran en encontrar objetos conspicuos captados en las imágenes, estos se basan en los contrastes que resaltan entre las diferentes regiones u objetos en la imagen; la detección de objetos sobresalientes da la opción mediante las semejanzas que produce contraste y ese contraste muestra la prominencia.

Primero se definen las semejanzas a las señales de prominencia o elevación un ejemplo de esto es cuando el ojo humano capta la imagen y el contraste entre la persona o el objeto con el fondo hace que el objeto sea destacado , segundo se codifican matrices de semejanza de contextos de contraste, en esta parte el mapa de prominencia de los pixeles hace probable que sea el objeto destacado, al transformarse el contraste múltiple se convierten las matrices de disimilitud en ultra contraste unificado, este ultra contraste asigna un valor a las prominencias y procede a la regresión logística.

Los algoritmos de detección de prominencia basados en contraste toman demasiado tiempo para clasificar la imagen, la detección de regiones salientes basada en contraste trata las diferencias de prominencia o colores parecidos como una sola región lo que restringe la selección de otros objetos. Se descubrieron características basadas en el aprendizaje profundo

en la técnica de detección de objetos salientes, esto hace complementario el experimento de características de ultra contraste. [24]

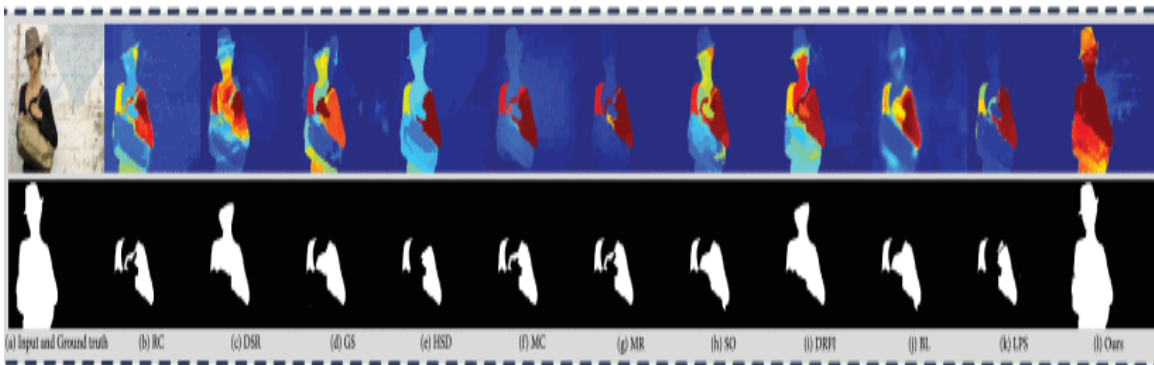


Ilustración 8

Autor: [24a]

o) Reconocimiento simultáneo de objetos 3D y estimación de pose basada en imágenes RGB-D

En este artículo se propone el uso del algoritmo de pose y reconocimiento de objetos RGB-D basado en nubes de puntos coloreadas es una función esencial. En el reconocimiento de objetos en la técnica de visión por computadora y sistemas de visión robótica, este tiene un funcionamiento donde las imágenes de color y su profundidad son captadas por el lente de la cámara RGB-D lo que origina una nube de puntos de color de la imagen, de esta forma el módulo de reconocimiento de objetos puede identificar el objeto principal o de su interés (object-of-interest OOI), y hace un cotejo entre las imágenes registradas en la base de datos de esta forma busca coincidencias con las características de otras imágenes simultáneamente se aplica un proceso de votación de Hough.

Este trabajo evidenció que los resultados experimentales validan el rendimiento propuesto utilizado por el algoritmo de votación Hough para filtrar los errores de coincidencia de características y esta estima la pose 3D del objeto, se insertó la etapa de estimación de pose RANdom SAMple Consensus (RANSAC) dando resultados exitosos para entrega de información precisa de la pose 3D del objeto respecto al lente de la cámara. [25]

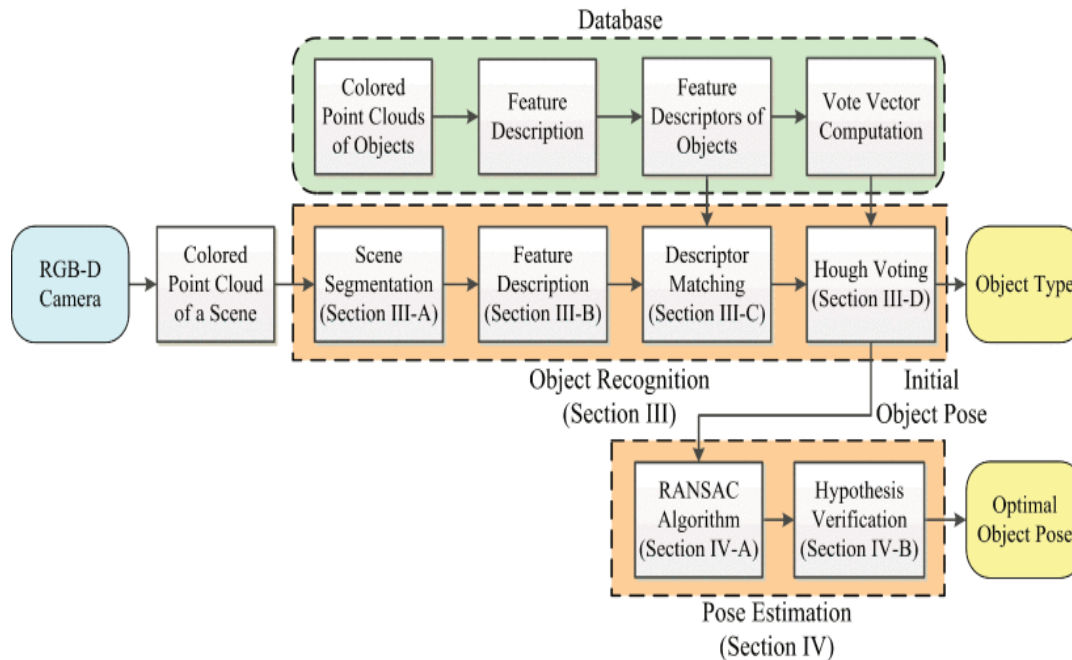


Ilustración 9

Autor:[25a]

p) *Aplicación de deep learning en robótica móvil para exploración y reconocimiento de objetos basados en imágenes.*

En esta investigación se empleó la técnica de aprendizaje Deep Learning donde el robot tiene como objetivo reconocer estructuras como escaleras, ascensores o corredores, establece movimientos para exploración con un láser mediante una cámara conectada a la web, la red neuronal convolucional 3D analiza los datos de la imagen donde el sensor crea patrones de la estructura y posibles movimientos de esta. En este punto el procesamiento del computador se implementa de forma eficiente al redirigir estos datos a las redes neuronales independientes en tiempo real, de esa manera explora paso a paso lo que se le enseña.

De esta manera pasaron el 3D Convolutional Neural Networks for Landing Zone Detection from LIDAR a zona boscosa calculando regiones adecuadas para el aterrizaje de helicópteros, el LIDAR obtiene la información en estructura tridimensional, se dividió el mapa obtenido en cuadrículas en subregiones. El proceso de aprendizaje se dio utilizando la librería Theano que hace cálculos sobre GPU acelerando el procesamiento del cálculo, planteando pruebas de mayor extensión, allí se incorporó mayor cantidad de modelos en 3d y objetos reales para entrenar efectivamente la red neuronal. [2] [52]

q) Reconocimiento de texturas en imágenes de proyectiles: un aporte a la identificación automática de armas

La siguiente investigación se enfoca en la “huella digital” de las armas, con el fin de determinar las similitudes de los rasgos característicos del patrón y el incriminatorio y establecer si provienen de la misma arma, el método utilizado para obtener esta información, es el de representación bidimensional de las regiones de interés, son las superficies de los proyectiles percutidos, que están dotados de características cuantitativas de las imágenes captadas, así se hace un soporte estadístico de diferencias entre las armas.

Creada por el Instituto Australiano de Seguridad y de Tecnología, el sistema llamado Fireball almacena imágenes de las estrías de proyectiles y vainillas de forma automática, haciendo una base de datos de características de espectros de las imágenes de las superficies de los mismos proyectiles y los dueños de las armas.

“permite almacenar y recuperar imágenes de la parte posterior de las vainillas y luego de forma interactiva, obtener la posición métrica de la impresión del percutor, la marca del eyector y la marca del extractor” [\[52\]](#)

r) Adquisición y visualización de escenas con foco variable utilizando procesamiento digital de imágenes

Depth of field (DOF) o profundidad de campo, este sistema óptico tiene configuración que referencia al rango espacial y de distancia, esta permite que la fotografía macro y microscópica se mantenga de forma nítida y enfocada. Funciona tras una secuencia de imágenes con focalización desde diferentes distancias, esto hace que al integrar las imágenes en una sola imagen que se muestra completamente enfocada.

El método general de FIMF fusión de imágenes multi-foco, este método se configura calibrando el sistema, continuado por la implementación de la técnica de fusión y termina con la visualización de dichas imágenes. Este sistema óptico es de bajo costo con campo variable en el rango 0,18 m.

El experimento realizado validó el método en escenas simples y complejas con dos escenas; la primera una escena con alto contraste en condiciones de laboratorio y la segunda una escena natural que presenta zonas con alto y bajo contraste.

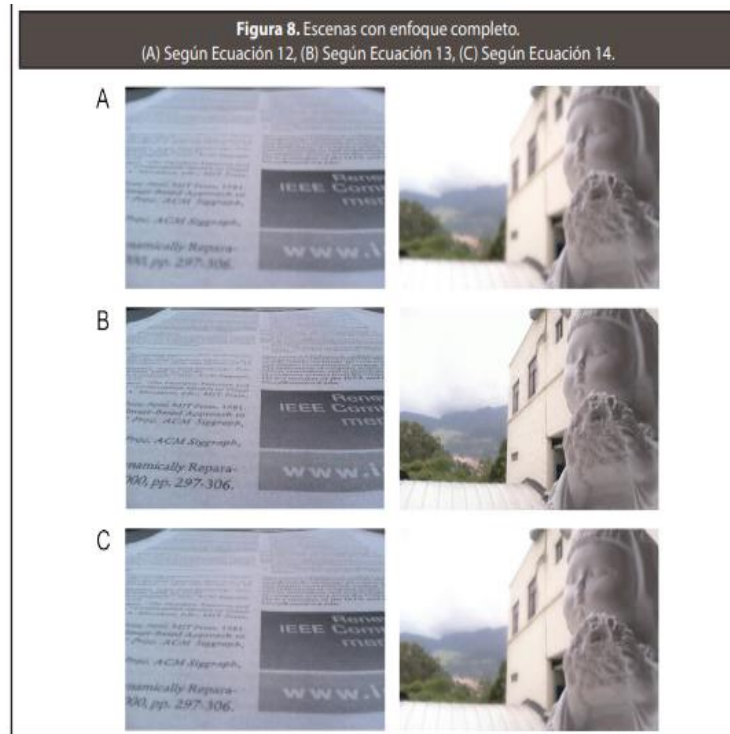


Ilustración 10

En la segunda imagen se demuestra el resultado de la obtención de imagen con foco completo para las escenas de la imagen anterior, resultado de utilizar la técnica del promedio de imágenes (A), el mapa de foco (B) y la combinación de ambas (C).

s) Detección de objetos complejos usando redes neurales y micro temblor ocular

Este estudio presenta un sistema reconocedor y seguidor en el cual dos redes neuronales conectadas en cascada, esta entrenada con criterios para procesar imágenes y producir reconocimiento confiable de un objeto que se mueve libremente visualizado por medio de una webcam. Estas redes neuronales se entrenan conociendo fenómenos subyacentes o sea mediante ejemplos que el sistema acoge recibiendo imágenes directamente mediante el movimiento del ojo. Este entrenamiento artificial se les llama fenómeno de micro temblor ocular OMT, características de los ojos animales, estas señales se guardan en una memoria de corta duración y son trasladadas a una segunda red neuronal que convoluciona a una única célula concepto que detecta el objeto seleccionado por la dirección ocular.

Este estudio propuesto estrechamente con OMT para el reconocimiento del objeto con aprendizaje progresivo el cual tuvo variaciones sobre el mismo objeto cómo: posición inclinación y escala captadas por la cámara web de manera eficiente manejando las condiciones dinámicas, que da como resultado la presencia del objeto seleccionado que es fácilmente

reconocible por la red neuronal con el sistema de seguimiento, para luego reconocerlo con una tasa de acierto final del 98% donde se comprueba que es posible descomponer el proceso de identificar un objeto móvil. [\[54\]](#)

t) Sistema de Reconocimiento de objetos en tiempo real

SURF (Speeded Up Robust Features)

Este algoritmo es de los más usados para la extracción de puntos de interés en el reconocimiento de imágenes, este detecta los posibles puntos de interés y la localización dentro de la imagen de una forma rápida. En esta investigación la universidad autónoma de Ciudad de Juárez en México. Realizaron los algoritmos de reconocimiento de patrones SURF Y SIFT adaptados a plataformas robóticas las cuales tienen tareas amplias como acomodar los libros de una biblioteca mediante selección y aislamiento. Por medio de la interfaz de programación de aplicaciones (API). Esta plataforma robótica móvil llamada de robotina tiene librerías de código libre open CV y sensor propio de visión del robot.

Sift(Scale-invariant feature transform)

Este algoritmo de reconocimiento de objetos se aplica en el área de visión artificial fue creado en 1999 por David Lowe, este algoritmo está encargado de extraer características en la escala de grises de las imágenes para luego reconocer dicha imagen dentro de las bases de datos incluso puede reconocer esta imagen dentro de muchas otras, aunque sus elementos estén en desorden las características son invariantes a factores de la escala donde puede haber cambios de iluminación y rotación. [\[55\]](#)

u) Scalable logo recognition in real-world images

Este artículo de investigación propone un reconocimiento de logotipos dentro de las imágenes este tiene el método de enfocar y codificar el diseño espacial de las características detectadas en el logotipo, estas características tienen estructuras espaciales básicas como bordes que llevan a una representación cuantificada de los puntos de interés del logotipo minimizando la detección de bordes falsos, propone una cascada de reconocimiento de múltiples clases de logotipos.

Consiste en que las imágenes de los logotipos sean etiquetadas manualmente, y son comparadas con las imágenes publicadas en Flickr, este conjunto de datos obtenido lleva a un conjunto de entrenamiento de datos de validaciones y pruebas donde se evalúa los puntos de referencia y diferentes puntos de vista para entrenar a la máquina efectivamente para el reconocimiento del logo con alta precisión. [\[56\]](#)



Ilustración 11

Esta imagen demuestra cómo se da la detección de puntos de interés del logotipo de Esso: imagen original 1, todos los bordes detectados de cualquier clase 2, reducción de bordes clase 3, triángulos detectados 4.

2.2 Marco Conceptual

Pixel: Superficie homogénea más pequeña de las que componen una imagen, que se define por su brillo y color. [4]

Imagen digital : la imagen digital es una representación numérica bidimensional, esta puede ser de imagen estática o dinámica, según la resolución es imagen matricial o gráfico vectorial. Esta es un producto del desarrollo de la tecnología que antecede a la fotografía, aborda una imagen desde que se toma por primera vez hasta procesos como filtración, compresión, optimización, impresión o almacenamiento en la nube. [5]

Procesamiento la imagen: procesamiento de la imagen, manipulación de la imagen, la muestra real de la imagen; de esta manera se pudo propiciar las tecnologías como la televisión, y todo aquello que transmite imágenes. [6]

Segmentación de imágenes: esta segmentación se puede dar por píxeles, abordando los puntos de la imagen se puede ver a que clase pertenece el píxel, estos atributos de la segmentación de imagen son: la luminancia en imágenes monocromáticas, los componentes de color en imágenes en color, textura, forma, etc. [7]

Realce o restauración de la imagen: “En la restauración de imágenes, se supone que el modelo matemático del proceso de degradación está disponible o puede derivarse con precisión examinando las imágenes de entrada. Los problemas de restauración de imágenes se pueden examinar en un marco matemático agradable. Sin embargo, aplicar estas técnicas para resolver problemas del mundo real es difícil” [8]

Interpretación de la imagen: se interpreta una imagen al extraer la información proveniente de esta donde se representa en dos o tres dimensiones, abarcando desde la fotografía a blanco y negro, pasando por imágenes satelitales hasta la resonancia magnética y nuevos desarrollos que están trayendo al siglo XXI la tercera dimensión mediante imágenes. [9]

2.3 Marco Teórico

2.3.1 Visión Artificial

La visión artificial o visión por computador, es la ciencia de la computación con el fin de procesar imágenes y en muchos casos entenderlas. [\[38\]](#)

La transformación de datos se da a través de fotogramas que dan una información que desenlaza una decisión, esto se da mediante una cadena de procesos que van desde la adquisición de la imagen, procesamiento, obtención de datos y ejecución de la acción predeterminada.

Los sistemas de visión artificial comprenden varios tipos de redes artificiales conectadas en cascadas, estas están entrenadas con diferentes tipos de aprendizaje los cuales se expondrán en el mapa conceptual de la próxima página.

La visión artificial el procesamiento de imágenes y la y el reconocimiento de imágenes, se han convertido en herramientas de alta utilidad dentro de los sistemas basados en conocimiento estos involucran frases como: procesamiento detección de bordes segmentación reconocimiento de localización e interpretación de la imagen.

El surgimiento de las computadoras se da en los años 50 al interior de la comunidad científica con el interrogante de si había la posibilidad de enseñar a la computadora de resolver problemas comprender lenguaje entre el ser humano y ellas o analizar información adquirida visualmente, orientando la emulación de la inteligencia humana hasta el día de hoy se han logrado grandes avances donde la visión artificial se ha asemejado al proceso de visión humana. Convirtiendo la matriz de puntos adquirida mediante una cámara en una imagen, donde se analiza la imagen mediante tareas adyacentes determinando características. También considera corregir vives diversos defectos cómo defectos en el sistema de adquisición, pérdida de datos durante la captura de la imagen, distorsiones en la escala o iluminación captada. [\[56\]](#)

Fases del proceso de funcionamiento de la visión artificial al determinar una imagen, fases no dependen una de otra ni tampoco son obligatorias en todas las técnicas que se desempeña en el momento.



Ilustración 12

Autor: [39]

a) Adquisición de la imagen:

Las propiedades de la captura, tipo cámara, mega pixeles, tipo lente, iluminación, etc. Es la digitalización de la imagen, quiere decir que pasa de lo físico a lo digital, como lo describe la siguiente imagen se necesita de una cámara y una tarjeta que adquiera la información.

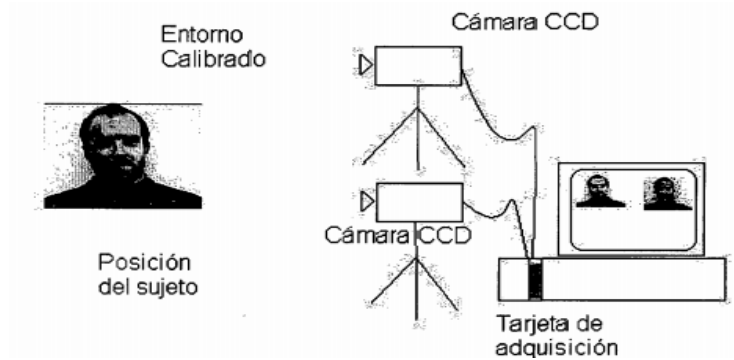


Ilustración 13 [47]

b) Procesamiento:

Reducción de lo que no es de interés para el entorno. Procesamiento el cual facilita etapas posteriores, eliminando partes de la imagen que no se toman en cuenta como:

1- *Amplitud de la escala*



Ilustración 14

También llamada como: amplitud tonal, rango tona, escala tonal, en las cámaras digitales para incrementar la resolución, esta es progresiva y tiene un límite en cada cámara dependiendo de lo programado en esta.

2- *Modificación de contraste*

El contraste se mide por la capacidad resolutive del sensor sobre el objeto. Este se acciona matemáticamente mediante la transformación de Fourier el cual presenta variables de frecuencia con una acción o contraste y fase.

Cuando la imagen es capturada tratamiento digital sobre la misma pasando por componentes como el sensor del objetivo la película, los cuales afectan la información original y por lo tanto hay pérdidas sobre esta. Las funciones del contraste son la rendición tonal o de frecuencia de respuesta espacial (MFT, modulation transfer function). Esta nos muestra las alteraciones sufridas por la imagen ya sea en variables de frecuencia y modulación dada por el paso de luz por cada componente del sistema de captura de imagen.

Entonces el contraste es la diferencia de tonalidades dentro de la imagen, estas tonalidades se perciben en líneas blancas y negras estas medidas se registran como líneas que no son puramente blancas o negras, sino que también se perciben como tonos grisáceos dadas las intensidades diferentes de la resolución y contraste del mismo objeto son dos caras de la misma moneda.

En esta imagen se puede evidenciar de forma más clara y nítida las tonalidades más altas de contraste a las más bajas esto quiere decir que a menor contraste es más clara la separación de líneas. [57]

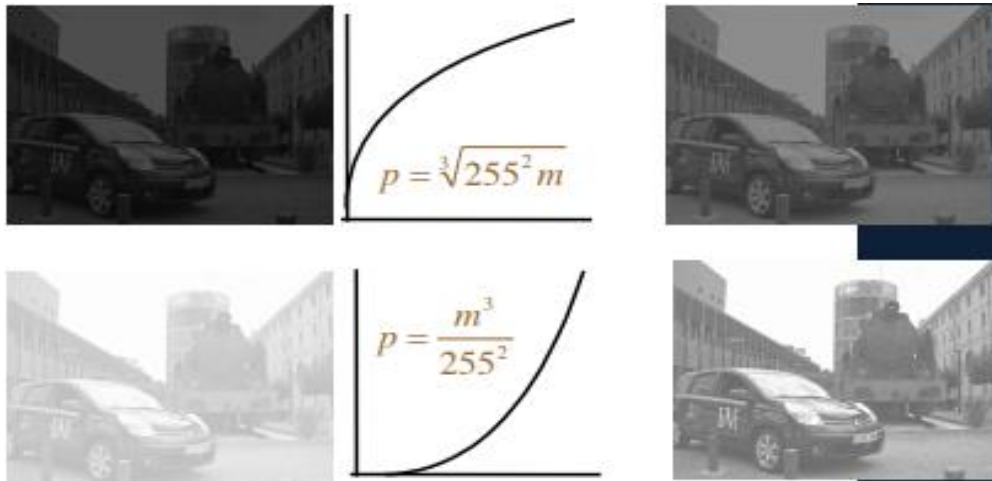


Ilustración 15

3- Modificación de histograma

El histograma es una gráfica que demuestra la distribución del conjunto de datos se presenta en barras y esta muestra la acumulación variabilidad o dispersión de los datos de el tema concreto este se usan para representar variables continuas ya sea cuantitativa o numérico. Cuando observamos una campana en el histograma este nos muestra la distribución simétrica estadística. [58]

La distribución de la campana será en tres calificaciones grupo a b y c:

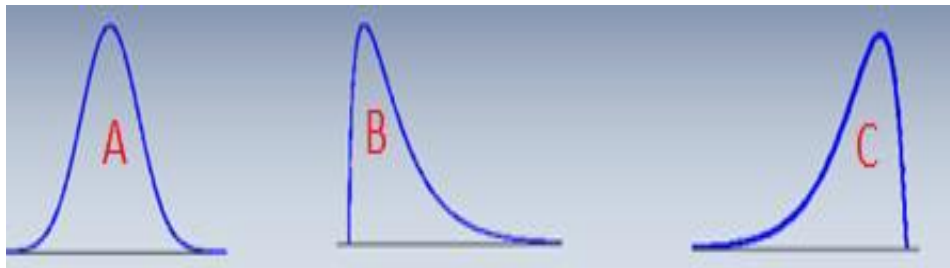


Ilustración 16

Autor: [58]

El grupo a muestra la distribución normal sobre los valores

El grupo b muestra que la distribución está sesgada a la derecha
La distribución de el grupo se muestra que está sesgada a la izquierda

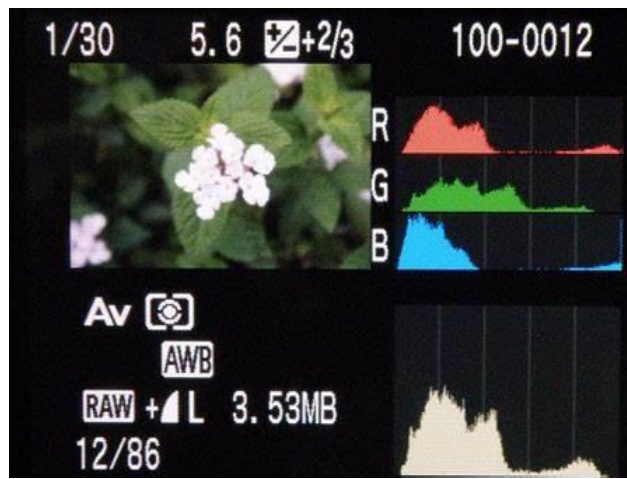


Ilustración 17

4- Reducción de ruido.



Ilustración 18

[\[48\]](#)

c) Deteccion de bordes:

Una de las etapas más importantes es la detección de bordes en este punto se define el fondo y el objeto y se separa el objeto de interés.



Ilustración 19

[49]

La detección de bordes es sumamente importante en el reconocimiento de imágenes, facilita la tarea de segmentación de regiones por ejemplo.

El Algoritmo de Canny propone tres criterios:

- Detección expresa: evita eliminar bordes importantes y no suministra bordes falsos.
- localización: muestra la distancia entre la posición del objeto y sus bordes.
- unidad de respuesta: integra las “respuestas múltiples correspondientes a un único borde”.

El algoritmo de Canny tiene tres pasos a seguir el primero es:

- calcular la magnitud y orientación del vector de cada píxel.
- Adelgazamiento de bordes derivados del gradiente.
- Histéresis de umbral: reduce la probabilidad de aparición de proximidades falsas. En la siguiente imagen se puede evidenciar como se obtiene mediante el algoritmo de Canny los bordes del objeto de interés. [59]

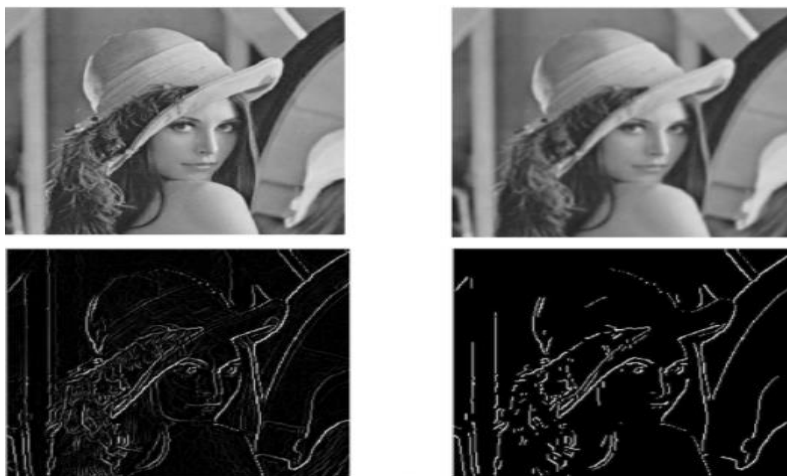


Ilustración 20

Autor: [\[59\]](#)

“La figura superior izquierda muestra la imagen original de Lena. La figura superior derecha muestra la imagen de Lena después de haber pasado un filtro gaussiano en el que se usó la máscara presentada en la figura 1a. La figura inferior izquierda muestra la imagen de Lena después de haber pasado por la supresión de no máximos. La figura inferior derecha muestra la imagen de Lena después del proceso de histeresis, y en los que se usó umbrales de 100 y 130.”

d) Segmentación:

El reconocimiento y extracción de los objetos presentes en la imagen. segmentación se seleccionan los píxeles con valores RGB (red, green and blue) o HSV (Hue, Saturation, Value) este modelo se basa en saturación y valores para combinar la colorimetría, de esta forma un objeto de diferente color se puede destacar.

“Split-and-merge”. El algoritmo segmentación

Existen otros modelos ampliamente usados en el procesamiento de imágenes, como pueden ser el HLS (posteriormente mejorado y llamado IHLS), esta herramienta detecta las curvas dentro de la imagen, busca características en los objetos, se centra en figuras geométricas como rectas, triángulos, objetos, círculos, etc.

La técnica de Hough está basada en modelos robustos frente al ruido y huecos en las fronteras del objeto, para aplicar esta técnica es necesario una imagen binaria. Un ejemplo de esto es la segmentación en umbral que tiene como objetivo encontrar puntos alrededor del objeto que estén alineados, esto da como resultado reducir los píxeles de una imagen para hacer la ecuación de la recta de coordenadas polares.

No que se interpreta como destacar los puntos de cruce de las curvas en el espacio de Hough, y la unión de las rectas pertenecen a dichos puntos en los espacios de la imagen. se hace de una forma muy sencilla es posible seleccionar los píxeles de interés en una recta para discriminar las características de los valores y se segmenta según cumpla los valores discriminatorios. [\[60\]](#)

En la siguiente imagen nos muestra cómo se detectan las rectas sobre el objeto, luego de esto se extienden las rectas para ser usadas con curvas las cuales pueden transformarse en formas paramétricas como triángulos elipses cuadrados etcétera. A éste se le conoce como la técnica generalizada de Hough:



Ilustración 21

Autor: [60]

e) Procesamiento

El algoritmo de procesamiento_FPGAs por medio de un spartan 3 puede realizar algoritmos de procesamiento de imágenes que corrige errores que requiere gran carga de procesamiento este se usa en los programas de lógica programable y procesamiento digital; este cuenta con gran diversidad de filtración de la imagen, lo que permite el realce y la obtención de imágenes de gran importancia como en el campo de la medicina. [61]

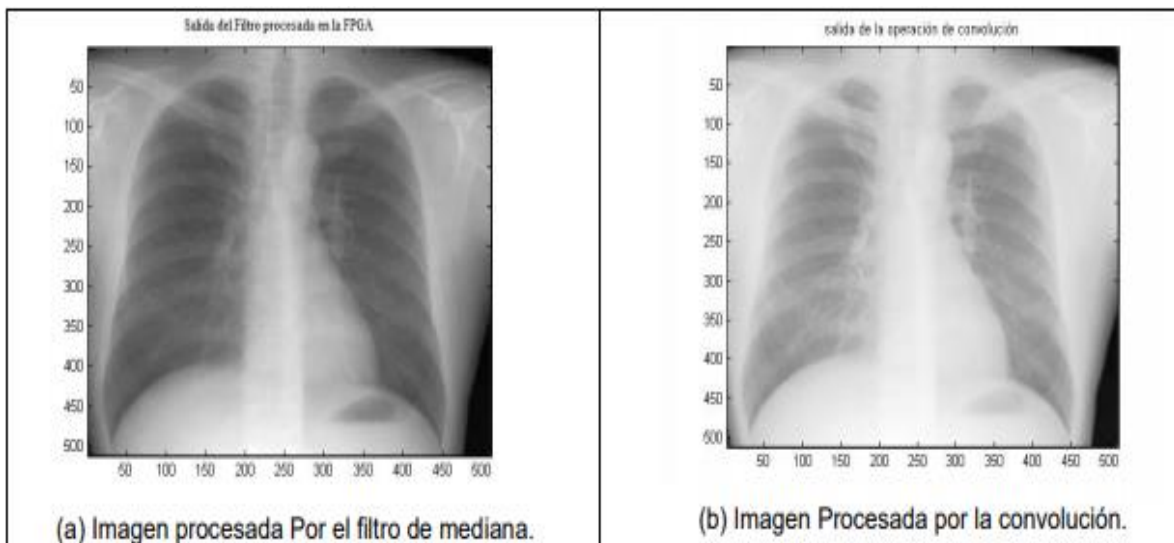


Figura 12. Resultados del Procesamiento en la FPGA.

Ilustración 22

- f) Extracción de características seleccionar características de identificación de los objetos que se desean.

Este subsistema se encarga de describir cada característica presente en la imagen, describe las características de los objetos en la imagen, lo que permite la identificación de modo cerrado; así que realiza dos tipos de descripciones, con el objetivo de identificar curvas, rectas, tipos de superficies, regiones, orientación, etc.

- g) Técnicas de reconocimiento de imagen:

Dar información al software para tomar la decisión a que categorías pertenece cada objeto.



Ilustración 23

[\[50\]](#)

Uno de los algoritmos mas usados para el Reconocimiento de imágenes es el algoritmo de k-means explicado a continuación:

El algoritmo k-means:

Este algoritmo de clasificación no supervisado, agrupa objetos en K grupos basándose en las características del mismo, se promedian los elementos de cada grupo estableciendo un centroide. Las principales ventajas de este procedimiento *k-means* es lo sencillo y rápido. Luego de decidir el valor de k entonces el resultado finalmente depende de la inicialización de los centroides.

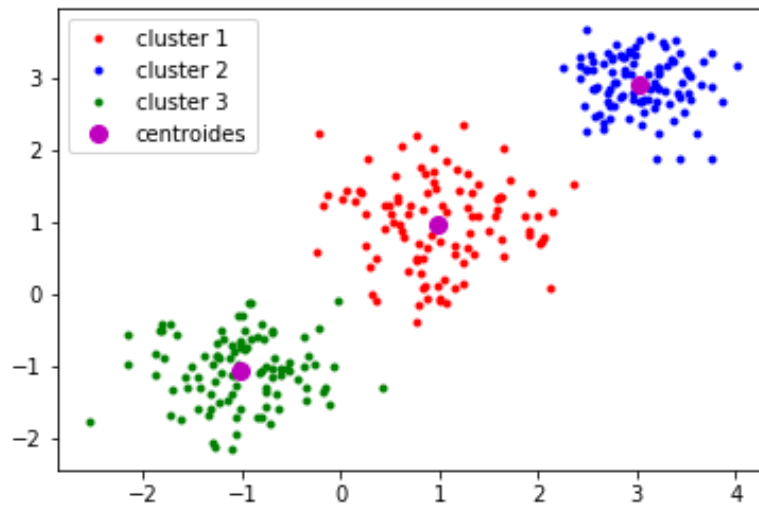


Ilustración 24

Datos aleatorios 2D para tres clusters

h) Clasifica el objeto y se entrena.

Este subsistema muestra las clases predefinidas dadas al objeto extraído de la imagen. Donde se observa la similitud de una imagen con otra mediante las características entregadas por la descripción del objeto, en algunos casos por medio de esta comparación de imágenes se construyen las estadísticas que continuamente configuran las redes neuronales artificiales. Estas brindan altos grados de complejidad a los problemas para clasificar y ubicar mejor las características para la toma de decisiones de la computadora.

Acción:

Este subsistema interactúa directa e indirectamente con el mundo físico puesto que entrega información obtenida de la imagen cruzada por el procesamiento hacia el usuario en el orden enseñado el cual se ejecuta dependiendo de los resultados obtenidos. Estos resultados se definen mediante el tipo de aprendizaje estipulado para que la función se desempeña en el sistema.

Un ejemplo de cómo se hace el procedimiento es de la siguiente manera:

- Captura: una distancia de la manzana a 60 cm del lente a resolución de 800 x 600, 24 bits.
- Preprocesamiento: enfocar solamente la manzana y desenfocar el fondo.
- Segmentación: reconocer y extraer la manzana de los objetos del fondo.
- Extracción de características: longitud, circularidad, color.

- Identificación del objeto: activación del algoritmo de clasificación dando como respuesta el reconocimiento de la fruta.
- Procesamiento digital de imágenes en drive.

La visión artificial es un sistema que posibilita la adquisición y análisis de imágenes de forma automática.

En esta imagen se describe como la visión artificial, un sistema que imita en lo posible la visión humana, puede implementarse a través de métodos de redes neuronales, aprendizaje estadístico, arboles de decisión, aprendizaje de variables ocultas, maquinas núcleo.

Estos métodos no son únicos, son cabeceras de otros algoritmos creados en base a estos con el fin de potenciar la visión artificial dependiendo de donde, como y cuando se necesita, no se utiliza el mismo algoritmo para visualizar una imagen desde el satélite a una imagen tomada desde un celular o por medio de patrones creados por láser, continuamente los tipos de aprendizaje como aprendizaje por observación, aprendizaje deductivo y conjunto de hipótesis tienen unos sub aprendizajes que se implementan para la necesidad del método que se utiliza.

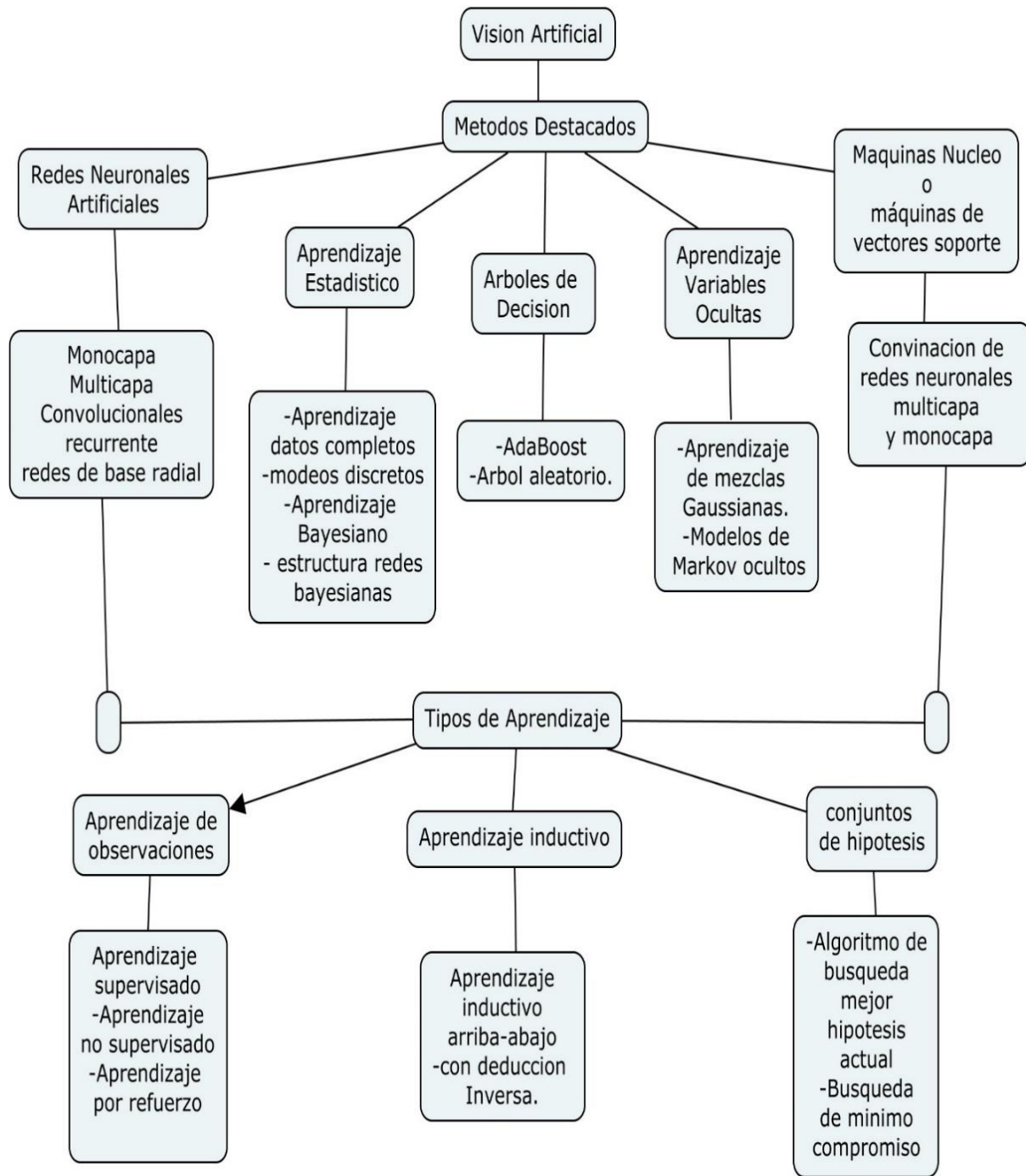


Ilustración 25

Autoría: Propia

Por medio de los siguientes métodos se puede implementar la visión artificial:

2.3.2 Redes Neuronales (CNN: neural network)

Son emuladores del comportamiento del cerebro humano que se caracteriza por aprender cuando obtiene una experiencia y extrae su conocimiento a partir de un conjunto de datos que logra analizar. Se utilizan para clasificar, predecir y controlar procesos. Las RNA son sistemas de información con capacidad de aprendizaje, esta de la mano con la estadística, utilizada para reconocimiento de patrones, procesamiento de imágenes, etc. [\[10\]](#)

2.3.2.1. Neurona Artificial.

Aprende: la información es adquirida por medio de datos, ejercicios o experiencia. Se le muestra un modo de entrada y ella misma se ajusta produciendo una salida sólida.

Generaliza: por naturaleza su estructura se generaliza automáticamente dando dentro de un margen establecido, respuestas correctas a entradas que puedan presentar pequeñas diferenciaciones.

Aísla: separa las características de los objetos.

Ventajas de la red Neuronal:

- Fácil inserción dentro de la tecnología existente.
- Aprendizaje Adaptativo
- Autoorganización.
- Tolerancia a fallos
- Operación en tiempo real. [\[29\]](#)

Entre los modelos de neurona tenemos: Asociador lineal, perceptron simple, redes Adeline, Madeline, red de hopfield, funciones de base radial, redes estocásticas. [\[11\]](#)

Las redes neuronales se componen de unidades de neurona con un peso numérico, que establece una conexión no lineal, donde la información pasa por las conexiones de entrada, función de entrada, función de activación, salida y finalmente por las conexiones de salidas.

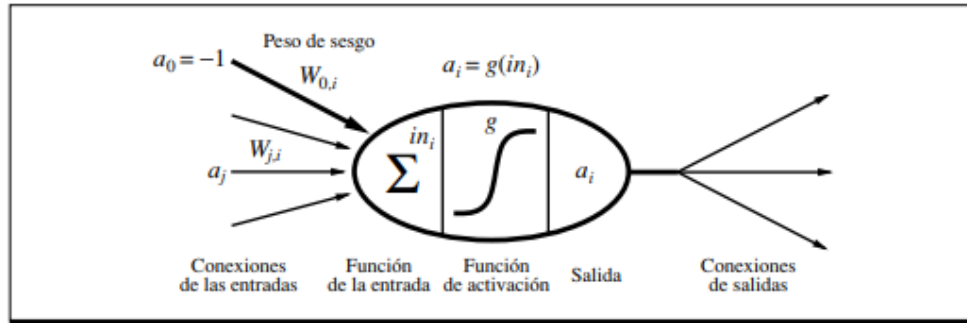


Ilustración 26

En la anterior figura podemos ver como los nodos se conectan a través de la unidad J que serían las conexiones de entradas hacia i la función de entrada, estas conexiones tienen un peso numérico W_{ji} estableciendo la fuerza, donde cada i calcula el ponderado de entradas. La función de activación g que tiene dos objetivos, la primera $a+1$ proporciona entradas correctas e inactiva la entrada a_0 cuando se dan entradas erróneas, la activación de la red neuronal no debe ser lineal puesto que colapsaría con esta sencilla activación. [40]

2.3.2.2. Tipo de redes neuronales

Las redes neuronales se clasifican en los siguientes tipos:

a) Redes neuronales monocapa

Son redes donde las entradas sólo pueden aprender fronteras lineales de decisión, estas se conectan directamente a la salida, no tienen el proceso de función de activación, también se le llama red perceptrón, estas unidades son independientes de las otras, algunos tipos de redes monocapa son: Hopfield, Brain state in a Box BSB, Learning Matrix. Este tipo de redes son muy utilizadas en circuitos eléctricos dado que sus matrices de diodo se usan como conexiones entre neuronas y se han implementado mediante hardware. [41]

b) Red Neuronal Convolutiva

La necesidad de mejorar los algoritmos computacionales enfocados en visión y reconocimiento de objetos, se da el desarrollo de una red neuronal convolutiva que trabaja con imágenes con resoluciones muy altas, de $1,980 \times 1,080$ píxeles, lo que conlleva un grave problema en las redes neuronales convencionales puesto que las capas ocultas conectadas con los píxeles de la imagen tienen una entrada por vez y se hace más lenta la neurona convencional y los tiempos de testeo pueden alargarse. [28]

c) Red neuronal convolucional multimodal.

Son estructuras de análisis de mayor profundidad lo que se traduce a multicapas, usadas para detallar complejas interacciones con el ser humano y la máquina, puede tener varias fuentes que aporte datos de forma simultánea. [27]

a) Red neuronal monocapa, b) red neuronal convolucional c) red neuronal recurrente

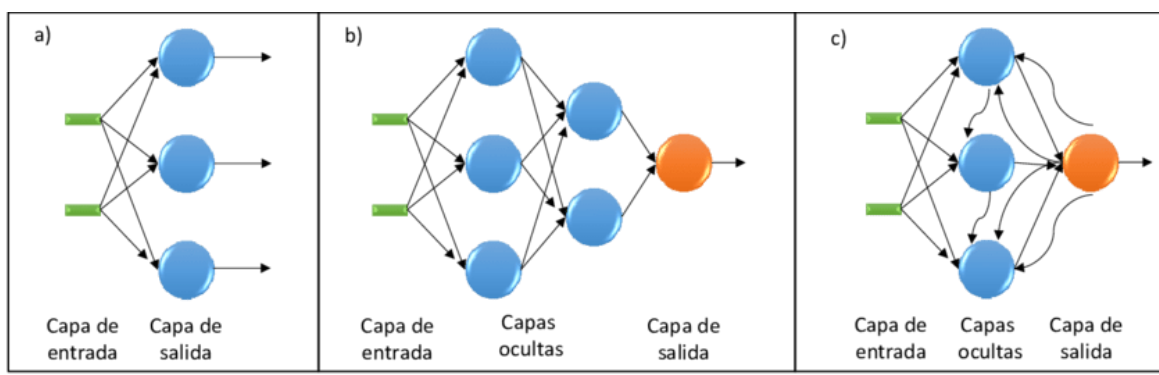


Ilustración 27

Autor: [42]

En la imagen a se observa la entrada y salida inmediata de la información, en la imagen b, la información entra, pasa por las capas ocultas donde guarda o adquiere información y en la imagen c la información entra, interactúa con los otros perceptrones para dirigir solo una información de salida.

d) Comparación entre red neuronal «tradicional» y red neuronal convolucional

En el siguiente cuadro se resumen las diferencias entre las redes neuronales tradicionales que se encargan de las características semánticas y las redes neuronales convolucionales encargadas de mejorar los algoritmos que soportaran los pesos de la metodología que requiera el programa.

Tabla 1

Características	Red «tradicional» Feedforward	Red Neuronal Convolutacional CNN
Datos de entrada en la Capa Inicial	Las características que analizamos. Por ejemplo: ancho, alto, grosor, etc.	Píxeles de una imagen. Si es color, serán 3 capas para rojo, verde, azul
Capas ocultas	Se elige una cantidad de neuronas para las capas ocultas.	Se tiene de tipo: * Convolución (con un tamaño de kernel y una cantidad de filtros) * Subsampling
Capa de Salida	La cantidad de neuronas que se quiere clasificar. Para «comprar» o «alquilar» serán 2 neuronas.	Debemos «aplanar» la última convolución con una (o más) capas de neuronas ocultas «tradicionales» y hacer una salida mediante SoftMax a la capa de salida que clasifica «perro» y «gato» serán 2 neuronas.
Aprendizaje	Supervisado	Supervisado
Interconexiones	Entre capas, todas las neuronas de una capa con la siguiente.	Son muchas menos conexiones necesarias, pues realmente los pesos que ajustamos serán los de los filtros/kernels que usamos.
Significado de la cantidad de capas ocultas	Realmente es algo desconocido y no representa algo en sí mismo.	Las capas ocultas son mapas de detección de características de la imagen y tienen jerarquía: primeras capas detectan líneas, luego curvas y formas cada vez más elaboradas.
Backpropagation	Se utiliza para ajustar los pesos de todas las interconexiones de las capas	Se utiliza para ajustar los pesos de los kernels.

Autor:[26]

2.3.3 Máquinas núcleo, máquinas de vectores soporte (SVMs)

Las máquinas núcleo suministran el mundo de las redes neuronales multicapa y monocapa, usando algoritmos de aprendizaje con funciones no lineales complejas, este se desarrolló como un método de clasificación binaria, pero tiempo después se empleó en situaciones como clasificación múltiple y regresión, está catalogado como un método de aprendizaje estadístico y machine Learning. [43]

Las máquinas de vectores de soporte forman una de las técnicas de clasificación supervisada más potentes, con clasificadores como LIBSVM, LIBLINEAR, el SVM tiene como objetivo concluir una frontera de decisión fuese lineal o no, hay un espacio de características de modo que los análisis posteriores que se clasifican de forma automática en uno de los dos grupos separados por el hiperplano. [51]

2.3.4 Aprendizaje estadístico

Es aprendizaje estadístico se da desde cálculos de medias y medianas hacia modelos complejos que se encargan de crear un conjunto de datos con la capacidad de predecir fenómenos en relación a lo aprendido. Estos datos estadísticos no son 100% exactos puesto que los datos dan una forma de razonamiento a partir de las variables e hipótesis apuntando a una probabilidad. [44]

a- Aprendizaje Bayesiano

La base de probabilidades bayesiana es subjetiva, está ligada al cambio acorde a la evidencia adquirida, por ende, la interpretación de los datos está ligado a hipótesis las cuales contrastar antes del estudio. El conjunto de probabilidades a priori de todas las hipótesis con factor de bayes permite una forma de calcular las probabilidades de la hipótesis donde se observa: promedios, correlación y asociación. [45]

b- Definición de red bayesiana:

La red bayesiana relaciona las variables aleatorias y las probabilidades donde los nodos representan variables aleatorias y los arcos representan relaciones de dependencia directa entre las variables, esto permite que se dé una decisión dándole solución al problema en casos de incertidumbre, en la siguiente imagen se puede evidenciar el modelo, donde es una variable aleatoria Z dependiente de dos factores F1 y F2.

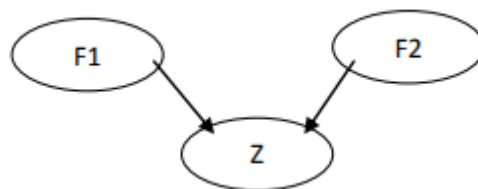


Ilustración 28

Autor: [46]

2.3.5 Árboles de decisión

a- Bosque aleatorio.

Es un algoritmo de personalización utiliza en la Personalización automatizada y en la Segmentación automática, donde se utilizan varios tipos de aprendizaje para tener mayor rendimiento de predicción.

Este árbol aleatorio de personalización automatizada y segmentación automática suelen integrarse con cientos de árboles de decisiones siempre con el fin de predecir más que lo que podría predecir un solo árbol. [\[31\]](#)

b- AdaBoost

Una técnica diseñada para combinar reglas.

Utiliza la combinación lineal de otras reglas sencillas, creando un clasificador robustecido con un numero bajo de errores respecto a los datos de entrenamiento.

En primer lugar, las reglas débiles aprenden en forma secuencial y distribuyen los pesos de datos de entrenamiento, durante la construcción de la combinación se puede dar que los pesos se actualizan mientras se adquieren nuevos datos o aumenta la dificultad del aprendizaje. [\[35\]](#)

2.3.6 Técnica de visión por computadora

Se da mediante la extracción automatizada de la información de una imagen ya sea posición de la cámara, búsqueda de contenido, reconocimiento de objetos, etc. Esta visión trata de emular la visión humana claro que con un enfoque estadístico y geométrico para resolver problemas. Le integran varios campos como la programación, modelado y matemática esto le convierte en un campo continuamente estudiado y actualizado. [\[37\]](#)

Fortalezas de técnicas.

Las técnicas presentadas en este cuadro son las expuestas por las investigaciones en el marco referencial de este documento, aquí se expone la técnica, el método o algoritmo al que pertenece y la fortaleza que presenta respecto a las áreas por las cuales debe pasar una imagen para tener el proceso de visión artificial y finalmente clasificación de la imagen.

Tabla 2

Técnicas	Método al que pertenece	Fortaleza
SLAM	Aprendizaje estadístico	Detección y Reconocimiento
Campo Aleatorio Condicional	Modelos Márkov Ocultos	Segmentación
Caracterización de Imágenes Binarias	Visión Por computadora	Detección de Características
Imágenes multiespectrales	Bosque Aleatorio	Reconocimiento
Procesamiento de imágenes morfológicas	Maquinas núcleo o máquinas de vectores soporte (SVMs)	Caracteriza y reconoce.
Detección de bordes	Aprendizaje estadístico	Detección de bordes de imagen
Técnica de imagen térmica	Aprendizaje automático	Adquisición de la imagen
Detección de objetos destacados	visión por computadora	Detecta y segmenta
Sistema de reconocimiento facial	Árbol AdaBoost	Reconocimiento +clasificación y procesamiento
Técnica data mining	Aprendizaje Estadístico	Detección + selección+ procesamiento+ transformación +patrones +interpretación
Red Convolutacional	Red Neuronal	Clasifica y segmenta
Red Convolutacional Multimodal	Red neuronal	Clasifica y discrimina
Mezcla Gaussiana	Aprendizaje de variables ocultas	Agrupación de datos

Autoría: Propia.

2.3.7 Reconocimiento de objetos

El aprendizaje supervisado proporciona una plataforma de estudio de reconocimiento de objetos, se le enseña mediante ejemplos positivos y negativos, esto significa por medio de rostros (positivos) y objetos (negativos) cada una de estas imágenes recibe una etiqueta. El reconocimiento de estas imágenes puede ser por diferentes técnicas bastante completas para abarcar un algoritmo que soporte el aumento de información, estas pueden ser: perceptrones multicapa, redes neuronales convolucionales, árboles de decisión, clasificadores por el vecino más cercano, y máquinas núcleo, primero se presenta la segmentación de la

imagen puesto que toda imagen que capte el visor puede captar múltiples objetos y estos se dividen en grupos de píxeles que forman cada objeto.

La imagen dividida por cada objeto se clasifica, se determina y se etiqueta.

Siguiente a esto se segmenta la imagen de arriba abajo (top-down) o abajo a arriba (bottom-up).

Aquí se busca un conjunto de píxeles que clasifican la cara o el objeto, la estrategia top-down es de alta complejidad computacional, dado que necesita gran agilidad para examinar múltiples ventanas con diferentes tamaños e instantáneamente comparar con otros objetos de características similares.

Seguido a la clasificación de estos píxeles se asegura el reconocimiento robusto del objeto, con variantes de iluminación, postura, luego de definir las características de valores de intensidad del píxel en sí mismo.



Ilustración 29

En la imagen anterior se utilizaron métodos estadísticos a partir de bases de datos de imágenes obtenidas de los cubos, por medio de estas redes neuronales y árboles de decisión con características determinadas por filtros de barra y borde, y modelos de Bayes sencillos con pequeñas características que muestra finalmente vectores de flujo comparando las 2 imágenes anteriores. [\[53\]](#)

Capítulo 3

Conclusiones

El reconocimiento de objetos ha tenido múltiples avances a medida que avanzan los desarrollos computacionales. Se describe de forma detallada el proceso requerido por la computadora para llevar la visualización de la imagen al tratamiento de esta, desde su captura, detección de bordes, procesamiento, reconocimiento y accionamiento de resultados, los cuales son enseñados a la maquina por medio de tipos de aprendizaje que continuamente se han desarrollado con el fin de ser partícipes de los diferentes tipos de técnicas y métodos empleados para que la computadora reconozca la imagen.

Existen diferentes métodos para el reconocimiento de imágenes, estos evalúan los rasgos enseñados por la persona que crea el código, y se pueden entrenar para buscar figuras geométricas. La captura de imágenes depende del dispositivo que se usa, entre los que hay cámara infrarroja que detecta mejor la figura de los barcos desde el aire o la térmica que detecta el aumento súbito de calor en la persona que dice mentiras.

Puede haber muchas técnicas que se emplean en el campo del reconocimiento artificial con el fin de procesar y reconocer la imagen hay una que sobresale gracias que esta se puede modificar continuamente y hay un robusto número de investigaciones que le respaldan. Las redes neuronales artificiales es uno de los métodos más usado en la mayoría de las investigaciones expuestas en este documento.

Se espera que este documento sirva de consulta a investigadores o en general a personas que requieran conocer las técnicas utilizadas en el reconocimiento de objetos.

Glosario

YOLO: You Only Look Once, «sólo se mira una vez»

SLAM Simultaneous Localization and Mapping

ORB-SLAM: es un optimizador de SLAM

BSIF: binarized statistical image features

CNN: neural networks Convolutional

HOG: Histograms of Oriented Gradient

MSI: imágenes multispectral

Arriba abajo (top-down)

Abajo a arriba (bottom-up)

Clustering: Algoritmo de agrupamiento.

Fotograma: Secuencia de imágenes captadas por cámaras de video.

Clouster: agrupar

FPGA: seguimiento al hardware en la matriz de puertas amigables en campo.

DOF: Depth of field

MFT: modulation transfer function

HSV: Hue, Saturation, Value

Clustering: Algoritmo de agrupamiento

Visión artificial: es una disciplina que incluye métodos para adquirir, procesar y comprender las imágenes, de esa manera se produce información en códigos o símbolos para ser interpretadas por medio del computador.

Inteligencia Artificial:

Es un campo científico de la informática el cual utiliza la combinación de algoritmos para crear programas y mecanismos que muestran comportamiento considerado inteligente al “imitar” capacidades del ser humano.

Deep Learning: es quien lleva los procesos de la red neuronal, mediante niveles de jerarquía, este envía la información de un nivel a otro

Referencias

- [1] Mercedes-Benz, «<https://www.mercedes-benz.com/en/innovation/autonomous/research-vehicle-f-015-luxury-in-motion/>,» 2020. [En línea]. Available: <https://www.mercedes-benz.com/en/innovation/autonomous/research-vehicle-f-015-luxury-in-motion/>.
- [2] J. Contrera-Rojas, Aplicacion Deep learning en robotica movil para exploracion y reconocimiento de objetos basados en Imagenes., Bogota: Universidad de los Andes., 2015.
- [3] J. M. García, «reconocimiento de texturas en imágenes de proyectiles: un aporte a la identificación automática de armas,» *CIENCIA E INGENIERÍA NEOGRANADINA*, 2012.
- [4] O. G. Chang, «Deteccion de Objetos complejos usando redes neuronales y micro temblor ocular,» *Revista facultad de ingenieria U.C.V*, 2013.
- [5] J. Estarita, A. Jimenez, J. Brochero y H. Escobar, «Sistema de Reconocimiento de objetos en tiempo real,» *Revista I+D en TIC*, 2016.
- [6] S. Romberg y L. Garcia, «Scalable logo recognition in real-world images,» *ACM International Conference on Multimedia*, p. 3, 2011.
- [7] J. J. Sanabria S., «Detección y análisis de movimiento usando visión artificial,» *Scientia Et Technica*, 2011.
- [8] E. C. Pardos, «Tecnicas de reconocimiento facial mediante redes neuronales.,» *Departamento de tecnologia fotonica. facultad de Informatica*, 2004.
- [9] E. Garcia y R. Osuna, «Fundamentos de la Fotografia,» *Universidad Nacional de Educación a Distancia*, 2009.
- [10] Universidad Nacional Autonoma de Mexico, « asesorias cuautitlan Histograma,» *Facultad de Estudios Superiores Cuautitlán*, 2019.
- [11] E. Alegre y G. Pajares, Conceptos y metodos en vision por computador, Comite Español de Automatica, 2016.

- [12] V. Alvear-Puerta, P. Rosero-Montalvo y D. Peluffo-Ordóñez, Internet de las Cosas y Visión Artificial, Funcionamiento y Aplicaciones: Revisión de Literatura, Quito: Universidad Técnica del Norte, 2017.
- [13] J. Valverde, «Detección de bordes mediante el algoritmo de Canny,» *Universidad Nacional de Trujillo*, 2007.
- [14] N. Aguirre- Dobernack, «Procesamiento de imagenes,» *Departamento de Ingeniería Electrónica.*, 2013.
- [15] S.-H. Yoon-Bustamante, «Algoritmos de procesamiento de imagen aplicados a la deteccion de figuras geometricas y sus propiedades.,» *PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO*, 2014.
- [16] J. Amat, «Maquinas de vector soporte,» *creative Coomons attribution Sharealike 4.0 international licence*.
- [17] S. Russell y P. Norvig, INTELIGENCIA ARTIFICIAL. UN ENFOQUE MODERNO. Segunda edición, Madrid: PEARSON, 2008.
- [18] J. W. Norton, Fisica, Massachusetts: Reverté, 2007.
- [19] J. J. Romero, C. Dafonte y Á. Gomez, Inteligencia Artificial y Computacion Avanzada, Fundacion Alfredo Cabañas Coleccion Informatica., 2007.
- [20] M. Dragoni y A. Azzini, SimBa: A novel similarity-based crossover for neuro-evolution, ScienceDirect, 2014.
- [21] J. Ramirez y M. Chacon, «Redes Neuronales Artificiales para procesamiento de imagenes, una revision de la ultima decada,» *Revista de Ingenieria Electronica*.
- [22] J. A. Ramírez Q y M. I. Chacón M, «Redes neuronales artificiales para el procesamiento de imágenes, una revisión de la última década,» *RIEE&C, REVISTA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y COMPUTACIÓN, Vol. 9 No. 1*, p. 10, 2011.
- [23] V. Vapnik, Statistics for engineering and Information Science, California: Springer, 1999.

- [24] B. Xue y N. Long, «Reconocimiento de objetos ISAR del mundo real y descubrimiento de relaciones mediante el aprendizaje de gráficos de relaciones profundas,» *IEEE*, pp. 43906 - 43914, 2019.
- [25] J. Jiang, G. Zou y M. Gao, «Reconocimiento de objetos basado en RGB-D mediante redes neuronales convolucionales multimodales: una encuesta,» *IEEE*, pp. 43110 - 43136, 2019.
- [26] Real Academia Española, «<https://dle.rae.es/p%C3%ADxel>,» Real Academia Española. [En línea].
- [27] P. Monteagudo Valdivia, «La imagen digital, una mirada interna.,» *Revista Cubana de Informatica Medica*.
- [28] G. Mota, F. Rizzo y J. Storer, «Compression of Hyperspectral Imagery,» *IEEE Xplore*, 2003.
- [29] L. Zhang, L. Wei, P. Shen y W. Wei , «Semantic SLAM Based on Object Detection and Improved Octomap,» *IEEE*, 2018.
- [30] J. Doyle y K. Bowyer, «Robust Detection of Textured Contact Lenses in Iris Recognition Using BSIF,» *IEEE*, 2015.
- [31] M. Zand, S. Doraisamy y A. Abdul, «Ontology-Based Semantic Image Segmentation Using Mixture Models and Multiple CRFs,» *IEEE*, 2016.
- [32] Y. Pan, X. Zhang y G. Cervone, «Detection of Asphalt Pavement Potholes and Cracks Based on the Unmanned Aerial Vehicle Multispectral Imagery,» *IEEE*, 2018.
- [33] W. Liu , M. Zhang y Z. Luo, «An Ensemble Deep Learning Method for Vehicle Type Classification on Visual Traffic Surveillance Sensors,» *IEEE*, 2017.
- [34] P. Zhao, H. Zhu, H. Li y T. Shibata, «A Directional-Edge-Based Real-Time Object Tracking System Employing Multiple Candidate-Location Generation,» *IEEE*, 2012.
- [35] B. Rajoub y R. Zwiggelaar, «Thermal Facial Analysis for Deception Detection,» *IEEE*, 2014.
- [36] . P. PEDRAZA, . O. BEATRIZ y P. RONDÓN , «Sistema de reconocimiento facial basado en imágenes con color,» *Revista UIS ingenierias*, 2011.

- [37] T. . W. Kiral Poon y M. R. Friesen, «Algoritmos para la detección de tamaño y color de imágenes de teléfonos inteligentes de heridas crónicas para aplicaciones sanitarias,» *IEEE*, pp. 1-10, 2015.
- [38] S. U. Sharma y D. J. Shah, «Un sistema práctico de detección de animales y prevención de colisiones que utiliza una técnica de visión por computadora,» *IEEE*, p. 1, 2016.
- [39] L. Tang, F. Meng, Q. Wu y N. Longdon, «Detección y segmentación de objetos destacados mediante ultracontraste,» *IEEE*, p. 2, 2018.
- [40] C.-Y. Tsai y S. H. Tsai, «Simultaneous 3D Object Recognition and Pose Estimation Based on RGB-D Images,» *IEEE*, p. 2, 2018.
- [41] J. D. Suárez, «Redes Neuronales Convolucionales,» Escuela Superior De ingeniería, Sevilla, 2017.
- [42] D. Match, «Redes Neuronales: Conceptos Basicos y Aplicaciones,» Universidad Tecnologica Nacional del Rosario, 2007.
- [43] X. B. Olabe, *Redes Neuronales Artificiales*, Bilbao: Dpto de Ingeniería de Sistemas Y Automatica, 2007.
- [44] O. J. Reyes-Ortiz, «TÉCNICAS DE INTELIGENCIA ARTIFICIAL UTILIZADAS EN EL PROCESAMIENTO DE IMÁGENES Y SU APLICACIÓN EN EL ANÁLISIS DE PAVIMENTOS,» *redalyc.org*, p. 10, 2019.
- [45] B. Rajoub y R. Zwigelaar, «Thermal Facial Analysis for Deception Detection,» *IEEE*, 2014.
- [46] R. Escudero, G. Escudero y S. Kanaan, *Inteligencia Artificial Avanzada*, Cataluña: Uiversidad Abierta de Cataluña.
- [47] L. Manjarrez, «Relaciones Neuronales Para Determinar la Atenuación del Valor de la Aceleración Máxima en Superficie de Sitios en Roca Para Zonas de Subducción,» *researchgate*, Junio 2014.
- [48] M. E. Rendón-Macías, A. Riojas-Garza y D. Contreras-Estrada, «Análisis bayesiano. Conceptos básicos y prácticos para su interpretación y uso,» *Revista alergia México*, 2018.

[49] M. R. L. «EL PAPEL DE LAS REDES BAYESIANAS EN LA TOMA DE DECISIONES,» *Laboratorio de modelamiento y simulacion*, 2011.