



Universidad Tecnológica De Panamá  
Facultad De Ingeniería De Sistemas Computacionales  
Lic. en Ingeniería de Sistemas y Computación



Memoria de Trabajo para la Asignatura:  
Tópicos I

Asignación N° 1 – Introducción a la Visión Artificial

Estudiante:  
Charles Chuez

Grupo:  
1IL141

Profesor:  
José Carlos Rangel Ortiz

2 Semestre, 2024

# Introducción a la Visión Artificial

La Inteligencia Artificial (IA), busca que los computadores realicen trabajos para los cuales, cuando un humano los realice se requiere inteligencia.

En la actualidad la IA como ciencia ha dado origen a diversas subramas de desarrollo, las cuales se enfocan en cumplir o lograr la misión general de la IA pero enfocada solo en uno de los aspectos de la inteligencia humana.

La Visión Artificial (VA) nace bajo este concepto y entre sus metas esta la emulación del comportamiento humano en lo relativo a la percepción visual o en otras palabras el sentido de la vista.

Desde su concepción se considera una de las entradas que más información puede proporcionar a un sistema.

Por ende, es de suma importancia el desarrollo de métodos que permitan extraer la mayor cantidad de información de este tipo de entrada.

La VA busca emular la visión humana y producir de una imagen de entrada, una salida de alto nivel, tal como lo haría una persona

Como ciencia se apoya en otras para lograr su cometido, a la vez que es utilizada por otras áreas para cumplir metas globales.

En la actualidad es una de las ramas de la IA que es más notada por las personas.

En el mundo cotidiano está presente en sistemas de vigilancia, cámaras fotográficas hasta en coches y robots autónomos, en los cuales cumple una función primordial.

La visión artificial, es un área de estudio multidisciplinaria, encargada de trabajar con imágenes (videos o imágenes estáticas).

Obteniendo de estas entradas, datos significativos para tomar decisiones, como las que realizan las aplicaciones tradicionales en las computadoras

Para la visión artificial o la implementación de este sentido en las computadoras, la entrada “cruda” de datos, se reduce a no más que a una matriz de números, representado de forma discreta algunos de los ejemplares de rayos de luz captados en un momento determinado.

Esta área, involucra varias técnicas y métodos para hacer posible que la computadora tenga, en cierto aspecto, un sentido de la vista similar al del ser humano.

En la actualidad la Visión por Computadora confía plenamente en algoritmos de Machine Learning y Deep Learning para la construcción de aplicaciones de visión artificial.

También se compone de varias tecnologías que trabajan de manera conjunta como lo son: Gráficos por Computadora, Procesamiento de Imágenes, Procesamiento de Señales, Tecnología de Sensores, Matemáticas e incluso Física.

La Visión Artificial, es la rama de la Inteligencia Artificial que se encarga del entendimiento de la información capturada por medios ópticos(cámaras) y almacenadas en un computador.

En la actualidad es ampliamente utilizada en conjunto con sistemas robóticos debido a que provee muchas soluciones que ayudan a lograr una mejor interacción con los usuarios.

En sistemas de robótica social se emplea el potencial combinado de la visión artificial y el Deep Learning para detectar las caras y emociones de las personas entre otras.

Es uno de los aspectos que mayormente ha contribuido a lograr una interacción más directa y fluida con las personas.

## Áreas de Investigación

### Reconocimiento de Rostros

El reconocimiento de rostros se inició aplicando técnicas de reconocimiento de patrones y biometría.

Este campo busca no solo la Localización de uno o varios rostros en la escena, sino también identificar a que persona pertenece dicho rostro.

En la mayoría de los casos se utilizan bases de datos donde se almacenan los rostros o características de los individuos que se desean reconocer.

Es en la actualidad uno de los campos que se ven comúnmente en aplicaciones de seguridad y teléfonos móviles.

## Detección y Reconocimiento de Objetos

Cuando se habla de reconocimiento de objetos, se trata del proceso en el cual un sistema inteligente trata de determinar si en una imagen dada se encuentra un objeto, generalmente estas imágenes ubican dichos objetos en el centro de esta o el objeto en cuestión ocupa un gran porcentaje de la imagen.

Sin embargo, no son las únicas formas de presentar la información a este tipo de sistemas.

En reconocimiento, por lo tanto, se habla de identificar si una imagen contiene un elemento en específico.

Cuando se habla de Detección, se pueden utilizar el mismo tipo de imágenes del caso anterior, pero la respuesta del sistema incluiría, las coordenadas de la ubicación del objeto dentro de la imagen, además de la categoría de este.

Por lo tanto, los procesos de detección involucran también un proceso de reconocimiento para conocer el tipo de objeto que se ha detectado en una escena.

## Segmentación Semántica

En esta área de investigación se busca hacer una clasificación de imágenes a nivel de píxel.

Se evalúa el contenido de la imagen y a cada uno de sus píxeles le será asignada una categoría o clase.

De esta manera cuando todos los puntos de la imagen han sido clasificados, es posible observar las zonas de la imagen en las que aparecen objetos, personas, estructuras, etc.

En la actualidad es un campo donde las soluciones predominantes son producidas utilizando algoritmos de Deep Learning.

## Estimación de Profundidad

La visión humana tiene la capacidad de percibir la profundidad o distancia a cuál estamos de las cosas, esto lo hace a partir de información en 2D.

La estimación de profundidad para las computadoras ha estado limitada al uso de equipos especiales como láser o de ultrasonidos.

Este campo se enfoca en las diferentes maneras en las cuales una computadora puede inferir como de lejos se encuentra algo, por lo tanto, involucra los procesos de creación de dispositivos de captura 3D, visión estéreo, cámaras 3D y el uso de algoritmos que permitan obtener información 3D a partir de una imagen o captura 2D.

## Reconocimiento de Texto

El reconocimiento de texto es uno de los campos de la Visión Artificial con mejores resultados y que ha estado presente de forma continua y desde hace varias décadas.

En este campo se siguen teniendo los mismos objetivos, pero con el pasar del tiempo ha ido variando la técnica o método que se emplea para lograr su fin.

Este campo busca descifrar el contenido de un texto y convertirlo a lenguaje editable en un procesador o en cualquier otro programa que lo utilice.

En la actualidad su foco de investigación está en la transcripción de la gran cantidad de manuscritos antiguos que existen almacenados en bibliotecas a nivel mundial, siendo estos los elementos con más dificultad al ser en gran porcentaje información manuscrita.

## Tratamiento de Imágenes Médicas

La medicina es uno de los grandes campos de aplicación para la Visión Artificial.

Existen diversas patologías que se identifican en una persona mediante la observación directa o el análisis de imágenes de ciertas partes del cuerpo.

Por lo tanto, es un campo donde se puede aplicar de muchas maneras.

Entre los problemas a los cuales se les está dando solución se pueden mencionar: determinación de calcificaciones en el corazón, detección de melanomas en la piel, análisis de radiografías y tomografías, entre otros.

## Conducción Autónoma

El campo de la conducción autónoma involucra muchas áreas de la inteligencia artificial, siendo la visión una de las que más aporta a la consecución de dicho fin.

En esta área los coches según el nivel cuentan con varios tipos de cámaras que permiten obtener información de los alrededores del auto y así tomar una decisión basada en los datos.

En este campo también entran en funcionamiento las diferentes cámaras que permiten captar información 3D del mundo.

## Entendimiento de Escenas/Reconocimiento de Lugares

El entendimiento de escenas se puede ver como el siguiente nivel de reconocimiento/clasificación de objetos.

Uno de los enfoques se orienta a determinar el contenido de la escena y según estos elementos determinar qué tipo de estancia se está viendo.

Otro enfoque se puede considerar el análisis global de la imagen, sin pasar por la identificación del contenido, para determinar el tipo de lugar correspondiente.

Es utilizado en aplicaciones de robótica móvil y conducción autónoma para lograr inferir la función que cumple un lugar tomando en cuenta su contenido.

## Información 3D

En línea de Visión Artificial, se trabaja con la información, estimada o capturada mediante los algoritmos de estimación de profundidad.

El objetivo de esta área es lograr la manipulación de información 3D y así utilizarla en procesos de reconocimiento y entendimiento de escenas.

De igual manera al trabajar con nubes de puntos de gran tamaño busca una forma de comprimir dichos datos asegurando la pérdida mínima de información en el proceso.

Es un área que presta mucho apoyo a procesos de grasping en robótica en la movilización de robots.

## Reconocimiento de Patrones

Es una de las áreas iniciales con las cuales se ha aplicado Visión Artificial.

Esta se enfoca en la detección de características sobresalientes en las imágenes y las cuales pueden ser repetibles con una cierta estructura en otra imagen.

De esta manera que cuando una imagen presente un conjunto de características en un orden o patrón, si este patrón se identifica en otra imagen diferente, se puede decir que tienen elementos en común y que pueden pertenecer a una misma categoría.

## Estimación de la Postura Humana

En este campo como su nombre lo dice, se enfoca en determinar la posición del cuerpo humano ya sea mediante el uso de una imagen estática proveniente de una foto o un frame de vídeo.

En los últimos años debido al resurgimiento del Deep Learning, ha obtenido mejores resultados y en tiempo real lo cual ha aumentado el número de aplicaciones de este tipo de solución.

## Restauración de Imágenes

En este campo se trata de la restauración automática de fotos y/o pinturas digitalizadas que han sufrido algún tipo de daño o percance debido a la acción humana.

Esta rama de la Visión Artificial ha permitido a estudiosos apreciar cómo podrían haber lucido algunas obras de la antigüedad y que han llegado con algún tipo de daño hasta nuestros días.

## Separación de Frutas

Dentro de las aplicaciones del reconocimiento y detección de objetos, esta su utilización en procesos de separación de elementos, basados en su apariencia visual.

En las últimas décadas se han desarrollado sistemas que permiten una identificación de frutas con algún tipo de desperfecto y de esa manera sacarla de la línea de producción de tal manera que se asegura una calidad en el producto entregado al consumidor.

Estos sistemas se caracterizan por trabajar a altas velocidades de tal manera que permite mantener un ritmo de producción sin amenazar el funcionamiento de la línea de producción ni el tiempo de vida del producto.

## Open CV

Es una librería open source de visión por computador, análisis de imagen y con algunas herramientas de aprendizaje automático.

Para ello dispone de infinidad de algoritmos que permiten, con sólo escribir unas pocas líneas de código, identificar rostros, reconocer objetos, clasificarlos, detectar movimientos de manos, etc.

Multiplataforma

Puede programarse con C, C++, Python, Java, Matlab, entre otros.

Se ha convertido en la librería por defecto para el desarrollo de aplicaciones con Visión Artificial.

## Visión Estereoscópica

Es un sistema que emplea dos cámaras/puntos de vista al momento de obtener información usualmente izquierda y derecha.

La visión humana funciona bajo este mecanismo.



El cerebro analiza las imágenes e identifica características similares en estas.

Posteriormente conociendo la separación entre los ojos es capaz de construir una imagen 3D del entorno.

Cada uno de los ojos acomoda el cristalino para enfocar correctamente un objeto, y convergen ambos ejes ópticos sobre el objeto que se mira.

Es cualquier técnica capaz de recoger información visual tridimensional y/o crear la ilusión de profundidad mediante una imagen estereográfica, un estereograma, o una imagen 3D (tridimensional).

La información tridimensional de profundidad puede ser reconstruida partir de dos imágenes usando una computadora para hacer relacionar los pixeles correspondientes en las imágenes izquierda y derecha.

Una variante del estéreo es el estéreo activo, en este enfoque además de la visión utilizando 2 cámaras se emplea un proyector que emite un láser o luz estructurada para facilitar el proceso de emparejamiento entre ambas imágenes.

## Iterative Closest Points ICP - Funcionamiento

Método iterativo que permite alinear dos superficies.

Se asume un solapamiento parcial (total en el caso ideal) y un conocimiento inicial de un alineamiento grueso.

Permite encontrar la transformación (en 3D) que mejor alinea los datos de entrada.

Inicialmente supondremos que las superficies a alinear vienen caracterizadas por conjuntos de puntos 3D.

Se utiliza una medida de distancia entre los elementos de un conjunto y los del otro. Inicialmente la distancia euclídea entre dos puntos.

# Aplicación Android para el reconocimiento de medicamentos mediante QR, texto a voz e inteligencia artificial

Issis Pitti | Roberto Aguilar | Juan Jose Saldana-Barrios | Yuraisma Moreno - Revista de Iniciación Científica, Vol. 10, No. 2, pp. 50 – 57, julio – diciembre 2024

**Objetivo:** Este artículo tiene como objetivo desarrollar una aplicación Android para el reconocimiento de medicamentos mediante códigos QR, texto a voz e inteligencia artificial con el interés de apoyar a las personas con problemas de baja visión que con frecuencia pueden enfrentar dificultades para identificar correctamente los medicamentos.

## Herramientas Utilizadas:

- Android studio
- Kotlin
- Text to speech (TTS)
- Código QR
- ZXing (zebra crossing)
- BudiYev
- Laptop Lenovo
- Dispositivo móvil con cámara
- Python
- Open AI
- Amazon Web Services (AWS)

**Aplicación de Visión Artificial:** El prototipo permitió generar un código QR con información relevante sobre un medicamento para luego ser escaneado, obteniendo, así como resultado, el texto introducido en el código QR y posteriormente convirtiéndolo a voz automáticamente.

**Resultados Obtenidos:** En una escala de 1 al 10, siendo 1 muy malo, no funcionó o no retornó nada y 10 muy bueno. Se obtuvieron resultados a la hora de escanear entre 8 y 10, siendo casi inmediato el escaneo, el retorno de audio es muy breve, entre 9 y 10 siempre y el texto de 10, siendo instantáneo el retorno, obteniendo resultados generales en todas las pruebas relativamente buenos.

# Sistema de visión artificial para gestión de calidad del Banano Cavendish en etapa de postcosecha

Brian O. Nieto | José Carlos Rangel - Revista de Iniciación Científica, Vol. 8, No. 2, pp. 32 – 42, julio – diciembre 2022 37

**Objetivo:** se busca desarrollar un sistema basado en visión artificial que permita determinar el estado de maduración de una fruta y estimar su tiempo de vida útil. Centrándose en este caso en el banano por ser una fruta de alta demanda a nivel nacional y que presenta diversos estados de maduración.

**Herramientas Utilizadas:** Para evitar que la recopilación de muestras se hiciera manualmente, desperdiciando tiempo y energía, es necesario construir un mecanismo a través del cual, utilizando tecnologías de IoT (Raspberry Pi 3 Model B, Arduino Nano), una cámara de ocho megapíxeles (Camera Module V2), un sensor de temperatura y humedad (DHT22) y un servicio de almacenamiento en la nube (MongoDB Atlas) se pudieran programar periódicamente capturas de imágenes en un entorno controlado durante todo el ciclo de maduración del Banano Cavendish.

**Aplicación de Visión Artificial:** Para cumplir con la fase de diseño se utilizó el lenguaje de programación Python, por su versatilidad y las librerías que posee para el manejo de datos. Entre estas librerías se encuentra OpenCV, que fue utilizada para limpiar las imágenes y generar el descriptor, como también Scikit-learn que proporcionó los algoritmos de Machine Learning para realizar la clasificación, así como las funciones para medir las métricas de rendimiento. Sumado a esto, el uso de Numpy permitió un mejor manejo de los arreglos n-dimensionales necesarios para realizar todo el proceso.

**Resultados Obtenidos:** Se realizaron múltiples rondas de recolección empezando a finales del mes de enero del 2020. Estas han tenido el objetivo de estudiar las propiedades ópticas que el Banano Cavendish adquiere a lo largo de su ciclo de maduración y a su vez capturar las condiciones ambientales (temperatura y humedad) en las que se tomaron las fotos. Además, se aplicaron otras variables como si el fruto está sólo o en un gajo, o si la caja de prueba tiene ventilación. Para manejo de posibles errores dentro del sistema de recolección de datos, de desencadenarse fallos durante las lecturas el sistema los agrega con un valor de cero. Estos posibles errores también fueron un factor a considerar al decantarse por tomar cinco lecturas diarias.