

Introducción a Visión Computacional

Eduardo Morales, Enrique Sucar

INAOE

Contenido

- 1 Introducción
- 2 *Visión Clásica*
 - Visión de Nivel Bajo
 - Visión de Nivel Intermedio
 - Visión de Alto Nivel
- 3 *Visión Moderna*
 - Redes Convolucionales
 - Transformers
- 4 Aplicaciones

Qué es Visión?

- “Visión es saber que hay y donde mediante la vista”(Aristóteles)
- “Visión es recuperar de la información de los sentidos propiedades válidas del mundo exterior”(Gibson)
- “Visión es un proceso que produce a partir de las imágenes del mundo una descripción que es útil para el observador y que no tiene información irrelevante (Marr)

Procesamiento de Imágenes

Entrada: imagen → Salida: imagen

- Remover defectos
- Remover problemas por movimiento o desenfoque
- Mejorar ciertas propiedades como color, contraste, estructura, etc.
- Agregar colores falsos a imágenes monocromáticas

Procesamiento de Imágenes

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

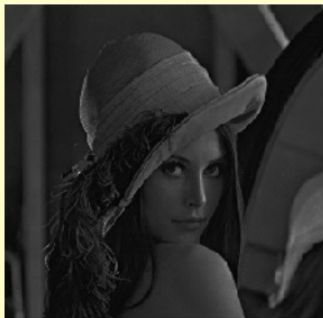
Visión de Alto Nivel

Visión
Moderna

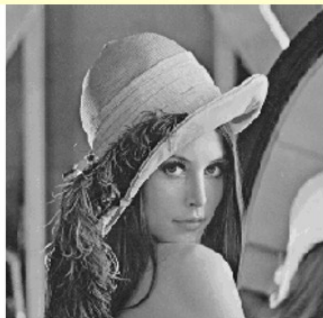
Redes
Convolucionales

Transformers

Aplicaciones



(a)



(b)

Visión Computacional

Entrada: imagen → Salida: descripción

- Determinar la identidad y localización de objetos en una imagen
- Construir una representación tridimensional de un objeto
- Analizar un objeto para determinar su calidad
- Descomponer una imagen u objeto en diferentes partes
- Seguir un objeto en movimiento
- Reconocer objetos específicos o clases de objetos
- Distinguir aspectos de un objeto (expresiones faciales)

Visión Computacional

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

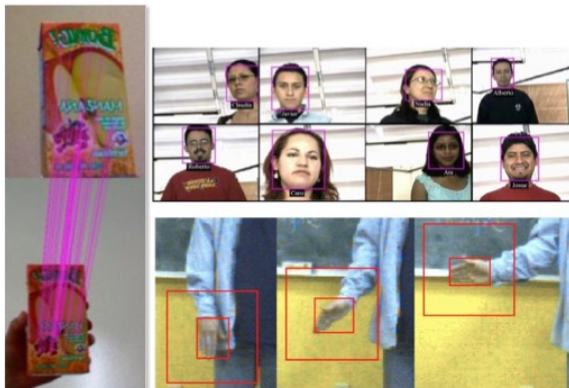
Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones



Visión Computacional

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales

Transformers

Aplicaciones



Niveles de Visión

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales

Transformers

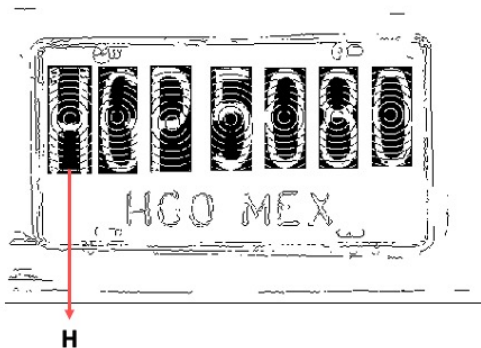
Aplicaciones

- Procesamiento de nivel bajo
- Procesamiento de nivel intermedio
- Procesamiento de nivel alto

Ejemplo - reconocimiento de placas de automóviles

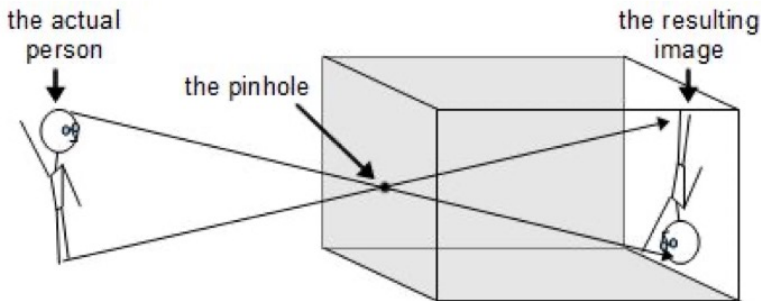
- 1 Captura de la imagen
- 2 Mejoramiento de la imagen
- 3 Binarización
- 4 Detección de orillas
- 5 Segmentación
- 6 Extracción de características
- 7 Reconocimiento

Ejemplo - reconocimiento de placas de automóviles



Formación de la imagen

- Modelo de cámara de “agujero de alfiler (pinhole)”



Representación

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

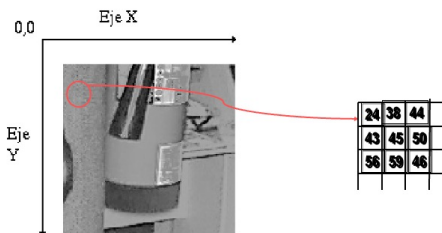
Redes
Convolucionales
Transformers

Aplicaciones

- Imagen monocromática: $I = f(x, y)$
- Imagen a color (RGB): $\mathbf{I} = [r(x, y), g(x, y), b(x, y)]$
- Imagen multi-espectral: $\mathbf{I} = [f_1(x, y), f_2(x, y), \dots, f_n(x, y)]$
- Imagen de rango: $I = Z(x, y)$

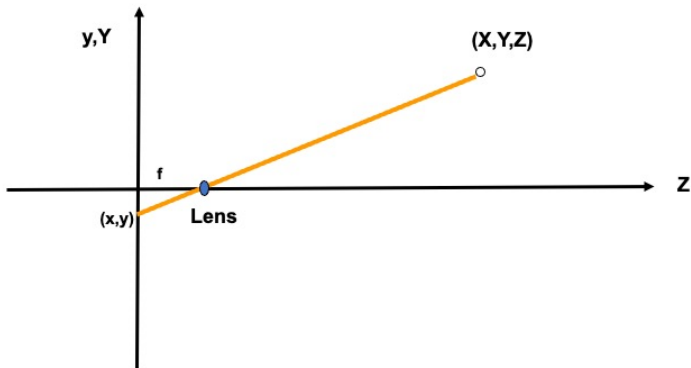
Digitalización de Imágenes

- Para almacenar y manipular una imagen en la computadora esta se digitaliza, tanto en valores (normalmente en 256 valores por canal) como espacialmente (muestreo de la imagen)
- Frecuencia de muestreo: > 2 veces mayor frecuencia (teorema de Shannon)
- Cada elemento de una imagen digital se le conoce como "pixel"(picture element)



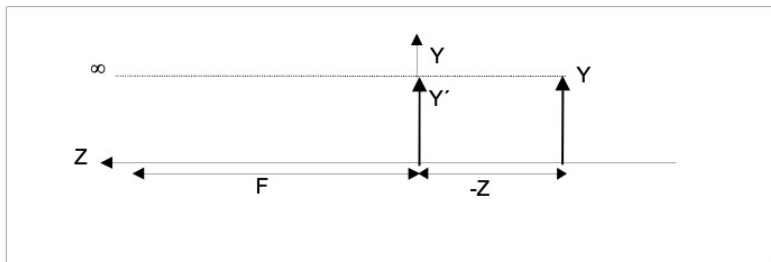
Proyección Persepctiva

- Modelo simplificado (cámara de agujero de alfiler) cuando los objetos están *cerca* de la cámara
- Por triángulos semejantes: $y/f = Y/(f - Z)$
- Por lo que: $y = fY/(f - Z)$
- En forma análoga: $x = fX/(f - Z)$



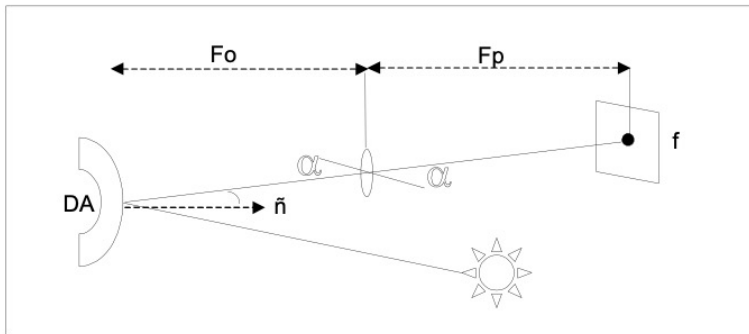
Proyección Ortográfica

- Modelo aproximado cuando los objetos se encuentran *lejos* de la cámara
- $x = X$
- $y = Y$



Reflectanica

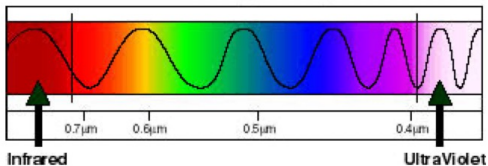
- La luz incidente en la cámara depende de tres factores:
 - 1 Fuente lumínica
 - 2 Geometría (ángulo entre el rayo incidente y la normal de la superficie)
 - 3 Propiedades del objeto



Color

- El *color* depende de la longitud de onda de la radiación electromagnética
- La luz visible está en el rango de 400 a 700 nm
- Nosotros (y las cámaras) percibimos el color por la combinación de 3 tipos de sensores que tienen mayor respuesta centrada a diferentes longitudes de onda – azul, verde y rojo

Visible Light Region
of the Electromagnetic Spectrum



Nivel Bajo

- Obtener características útiles para los siguientes niveles de visión:
 - Orillas o bordes,
 - Color,
 - Textura,
 - Gradiente,
 - Profundidad.

Mejoramiento de la Imagen

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones

- Previo a obtener características: resaltar aspectos deseados, eliminar ruido, mejorar contraste, etc.
- Técnicas de pre-procesamiento:
 - 1 operaciones puntuales,
 - 2 ecualización por histograma,
 - 3 filtrado.

Binarización

- Transformación no lineal en el que la imagen de salida tiene 2 valores: blanco (255) y negro (0)
- $S = 0$ si $E < T$
- $S = 255$ si $E > T$
- T es un umbral



(a)



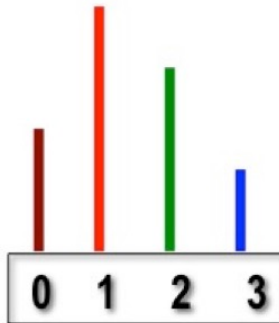
(b)



Histograma de Intensidades

- Distribución de niveles de intensidad en una imagen
- Provee un estimado de probabilidad de ocurrencia de cada nivel de intensidad: $p = n_r / N_t$
- Ejemplo:

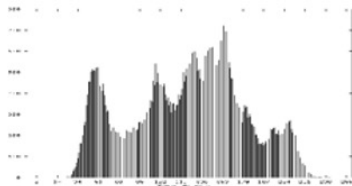
0	1	1	2
2	1	0	2
1	2	2	1
3	0	1	3



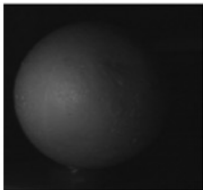
Ecuación por histograma – imagen original



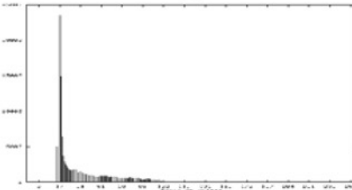
(a)



(b)



(c)

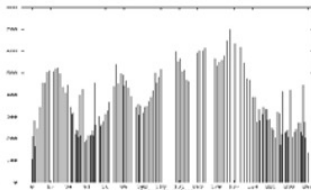


(d)

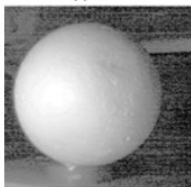
Ecuación por histograma – imagen ecualizada



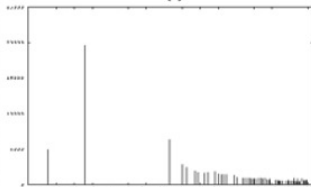
(a)



(b)



(c)

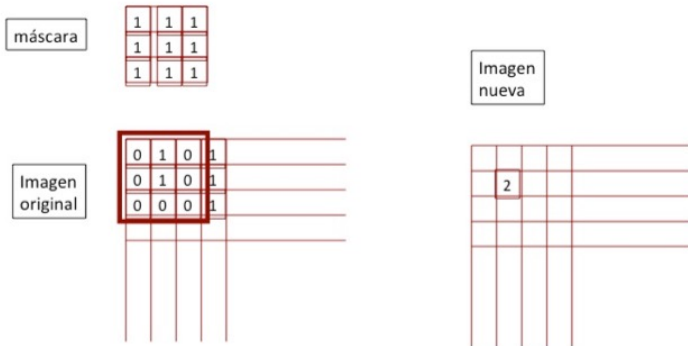


(d)

Filtrado

- Filtrar una imagen consiste en aplicar una transformación de forma que se acentúen o disminuyan ciertos aspectos
- Filtro: $g(x, y) = T[f(x, y)]$
- Tipos de Filtros:
 - Dominio espacial - convolución:
 $g(x, y) = h(x, y) * f(x, y)$
 - Dominio de la frecuencia - transformadas de Fourier y multiplicación: $G(u, v) = H(u, v)F(u, v)$
 - Diferentes filtros (máscaras) producen diferentes efectos en la imagen: suavizamiento, énfasis de altas frecuencias, etc.

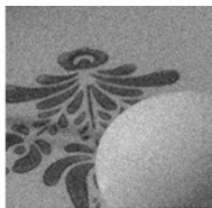
Filtrado en dominio espacial – convolución



Ejemplo de filtros de suavizamiento (pasa bajos)



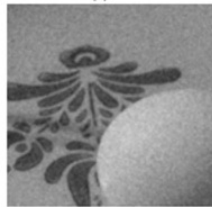
(a)



(b)



(c)

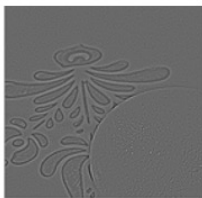


(d)

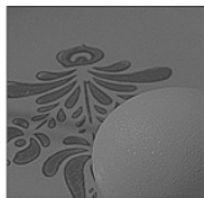
Ejemplo de filtros de énfasis de altas frecuencias (pasa altos)



(a)



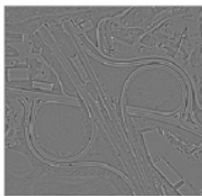
(b)



(c)



(d)



(e)



(f)

Detección de Orillas

- Las orillas de una imagen normalmente se detectan como segmentos o secciones de un borde (cambios bruscos de intensidad en la imagen) que se integran en etapas posteriores
- Las orillas son elementos importantes para etapas posteriores ya que normalmente significan los contornos de objetos o partes de objetos
- Incluso se ha encontrado que la visión biológica (primates, humanos) tiene neuronas dedicadas a la detección de orillas

Gradiente

- Una orilla en una imagen se presenta cuando la función de intensidad presenta un cambio fuerte localmente - tiene un valor alto de su primera derivada o gradiente
- Por lo que una forma de detectar orillas es aproximar el gradiente mediante la diferencia de valores contiguos en la imagen:
 - $df/dx = I_{1,2} - I_{1,1}$
 - $df/dy = I_{2,1} - I_{1,1}$

Operador de Sobel

- Un problema al aplicar el gradiente es que se amplifica el ruido
- El detector de Sobel se puede ver como la combinación de un filtro de suavizamiento con un operador de gradiente
- $Sobel = DG^T$
- Donde: $D = [-1, 0, 1]$, $G = [1, 2, 1]$

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Ejemplos de aplicación de Sobel



(a)



(b)



(c)

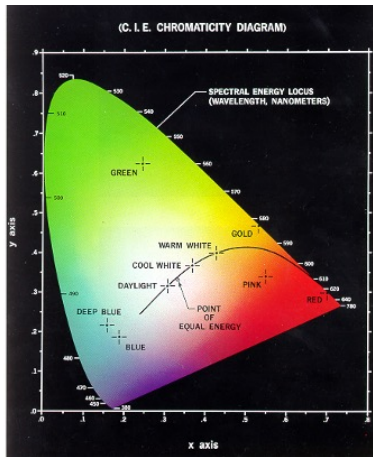


(d)

Color

- El color tiene que ver con la longitud de onda de la luz emitida o reflejada por un objeto
- Se puede ver como la combinación de 3 colores (primarios): rojo, verde y azul
- El diagrama cromático CIE es un estándar de representación del color que considera los siguientes colores primarios: azul, 435.8 nm; verde, 541.6 nm; rojo, 700 nm

CIE



Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo
Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

Aplicaciones

Modelos de color

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones

- Modelos sensoriales - orientados a los equipos: RGB, CMY, YIQ
- Modelos perceptuales - se asemejan a la percepción humana y se orientan al procesamiento de imágenes y visión: HSV, HLS, HSI

Modelo RGB



(a)



(b)



(c)



(d)



(e)

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

Aplicaciones

Modelo HSI

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

Aplicaciones



(a)



(b)



(c)

Ecuación en color

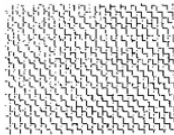
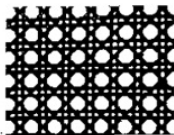
- Si aplicamos en cada componente en RGB, se altera el “color” de la imagen
- Una alternativa es convertir al modelo HSI y ecualizar sólo la componente de I - de esta manera se mejora el contraste sin alterar los “colores”



Textura

- Los objetos en general no son uniformes, tienen cierta “textura”
- Textura: “compuesta de pequeños elementos indistinguibles y entrelazados”
- La textura depende de la resolución
- La información de textura se puede usar para segmentación, reconocimiento de objetos y obtención de forma

Ejemplos de Textura



Descripción de Texturas

- Existen diferentes formas de describir una textura:
 - 1 Modelos estructurales
 - 2 Modelos estadísticos
 - 3 Modelos espectrales
- Un forma de caracterizar una textura es obteniendo el histograma de niveles de gris
- Del histograma se pueden calcular ciertos parámetros que lo caracterizan - una posibilidad es calcular sus momentos

Ejemplos de Histogramas de Texturas

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

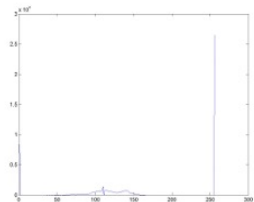
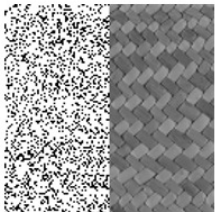
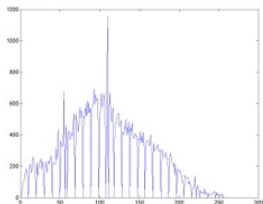
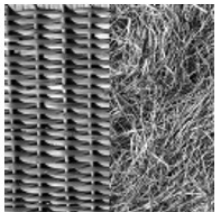
Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones



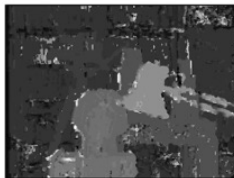
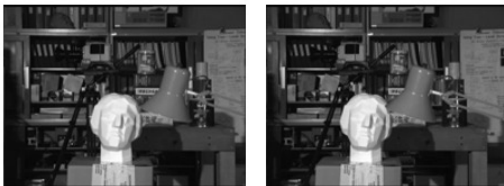
Visión Estereo

- Una forma de recuperar la tercera dimensión es mediante el uso de 2 cámaras o visión estereo
- La idea es combinar las imágenes de las 2 (o más) cámaras y mediante geometría obtener la profundidad de cada punto en la imagen
- Las 2 cámaras están en posiciones distintas a una distancia conocida

Algoritmo básico

- 1 Obtener dos (o más) imágenes
- 2 Identificar puntos correspondientes en ambas imágenes
- 3 Mediante geometría estimar la profundidad (Z) de cada punto - en base a la distancia entre los puntos correspondientes (disparidad) en las imágenes y los parámetros de las cámaras (longitud focal, distancia)

Ejemplo de Estereo



Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

Aplicaciones

Visión de Nivel Intermedio

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales

Transformers

Aplicaciones

- Agrupar las características obtenidas en el nivel bajo
- Generar una representación más compacta
- Aislar bordes o contornos (agrupar orillas)
- Obtener regiones (agrupar secciones uniformes)

Agrupamiento de Orillas

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales

Transformers

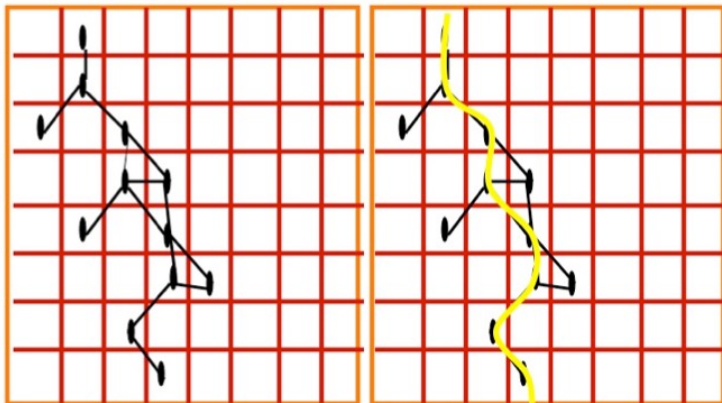
Aplicaciones

- Las orillas obtenidas mediante los operadores de detección de orillas se agrupan en bordes o contornos
- Técnicas principales:
 - transformada de Hough,
 - búsqueda en grafos,
 - agrupamiento perceptual.

Técnicas de Búsqueda

- Se considera a las orillas (magnitud, dirección) como un grafo
- El agrupamiento de orillas se basa en la búsqueda de una trayectoria en el grafo
- Algoritmo:
 - 1 Obtener magnitud y dirección de las orillas
 - 2 Convertir cada orilla en un nodo del grafo: peso del nodo = magnitud de la orilla; arcos = función de la dirección de la orilla
 - 3 Partir de una orilla (inicio) y buscar una trayectoria en el grafo que una las orillas
 - 4 La trayectoria corresponde al contorno

Grafo de Orillas – Contorno



Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales

Transformers

Aplicaciones

Segmentación

- Segmentación es separar una imagen en unidades significativas
- Dos alternativas básicas: encontrar bordes, determinar las regiones
- Las regiones se determinan en base a criterios de *homogeneidad*, en base a atributos como intensidad, color, textura, profundidad
- Dificultades:
 - variaciones de intensidad (u otro atributo) por la naturaleza de los objetos y condiciones de iluminación,
 - posible confusión de objetos similares,
 - otros artefactos como sombras, oclusiones, etc.

Técnicas de Segmentación

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales

Transformers

Aplicaciones

- Locales - agrupan pixels en base a sus atributos
- Globales - dividen la imagen en base a propiedades globales
- División-agrupamiento - combinan propiedades locales y globales

Método basado en Quadrees

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales

Transformers

Aplicaciones

- Se considera una región de interés de nivel de intensidad I
- Medidas de homogeneidad en base a promedio (μ) y desviación estándar (σ):
 - región uniforme en rango: μ cerca I y σ baja
 - región uniforme fuera de rango: μ lejos I y σ baja
 - región no uniforme: σ alta

Algoritmo

- Dividir imagen en 4, calcular (μ) y (σ) en cada cuadrante
- Si es región uniforme en rango, tomar como región base
- Si es región uniforme fuera de rango, desechar
- Si es región no uniforme, dividir en 4 y repetir pasos anteriores
- Tomar como región base la mayor y unir cuadrantes similares a diferentes niveles

Ejemplo de segmentación mediante QuadTree

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

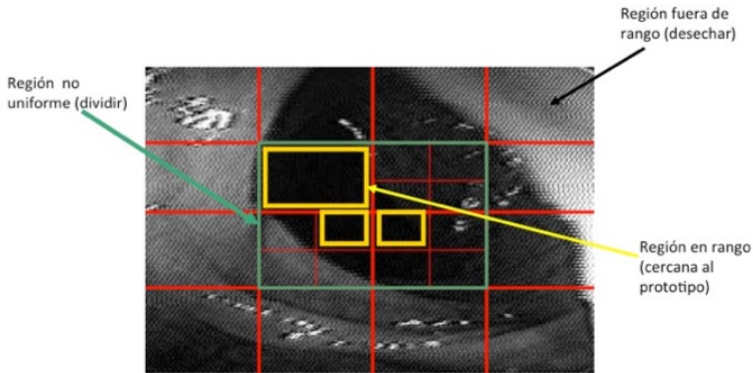
Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones



Visión de Alto Nivel

- Obtiene una interpretación consistente de las características obtenidas en visión de nivel bajo e intermedio
- Se basa en utilizar conocimiento de los objetos del dominio de interés
- En base al conocimiento y las características se realiza el reconocimiento
- Aspectos básicos:
 - Representación - forma de modelar el mundo, en particular los objetos de interés para el sistema
 - Reconocimiento - como el modelo y la descripción de la imagen(es) son utilizadas para identificar los objetos

Principales tipos de reconocimiento

- Basado en modelos
- Reconocimiento estadístico

Visión basada en Modelos

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales

Transformers

Aplicaciones

- Se basan en el uso de modelos geométricos predefinidos para cada objeto
- Tres componentes principales:
 - 1 extracción de características,
 - 2 modelado,
 - 3 apareamiento geométrico (matching).

Modelos en 2 dimensiones

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

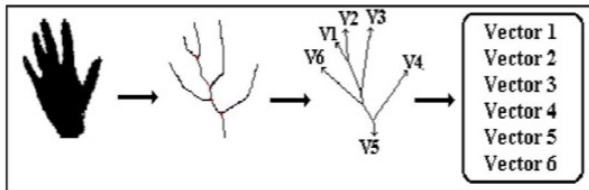
Redes
Convolucionales

Transformers

Aplicaciones

- En base a Contornos:
 - poli-líneas,
 - códigos de cadena,
 - descriptores de Fourier,
 - secciones cónicas.
- En base a Regiones:
 - arreglos de ocupación espacial,
 - eje "Y",
 - árboles cuaternarios,
 - esqueletos.

Ejemplos – modelos en 2D



Modelos en 3 Dimensiones

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones

- Representación de objetos en 3-D independiente del punto de vista
- Dos enfoques principales: (i) en base a su superficie, (ii) en base a una representación volumétrica
- Tipos de modelos:
 - Poliedros planos
 - Cilindros generalizados
 - Geometría solida constructiva

Ejemplos – modelos en 3D

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

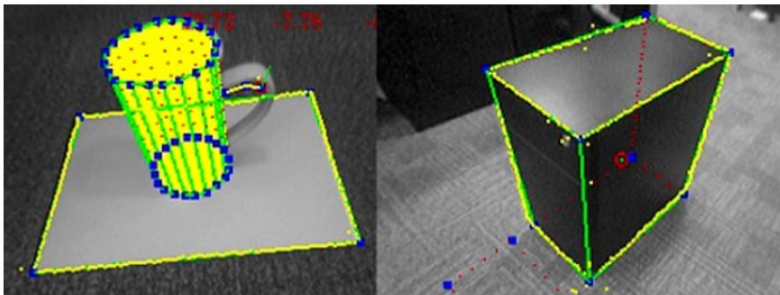
Visión de Nivel Bajo
Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

Aplicaciones



Reconocimiento estadístico

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo
Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

Aplicaciones

- Se utiliza un esquema de clasificación basado en técnicas estadísticas y de aprendizaje de máquina.
- Dos aspectos principales:
 - 1 Tipo de representación – características globales y/o locales
 - 2 Método de clasificación

Propiedades Globales

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo
Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

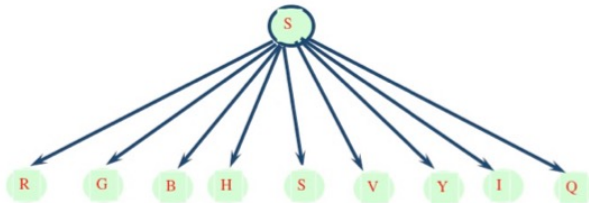
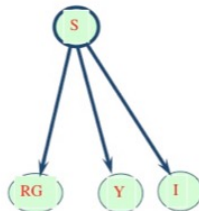
Aplicaciones

- Parten de una segmentación de los objetos de interés, la cual puede ser aproximada (rectángulos / ventanas deslizantes)
- Tres principales tipos de características:
 - Color
 - Textura
 - Forma

Color

- Caracterizan la distribución del color de una clase de objetos
- Se consideran diferentes modelos de color: RGB, HSI, YIQ, Luv, É o una combinación de estos
- Se puede utilizar directamente el histograma de cada componente del modelo de color, o una representación compacta de este (media, varianza, momentos, É)

Ejemplo – clasificación de piel en base a color



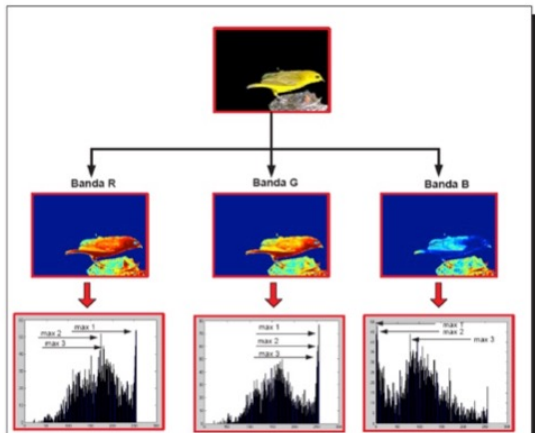
Textura

- Se busca caracterizar a la textura de un objeto/región con pocos parámetros, utilizando alguna de las técnicas como momentos del histograma, matrices de dependencia espacial, etc.
- Otra forma de caracterizar una textura es mediante la respuesta a un conjunto de filtros a diferentes escalas, como los filtros de Gabor

Forma

- Caracterizan la forma geométrica del objeto (región), normalmente en 2-D
- Para ello se pueden utilizar diferentes alternativas para representar la forma del objeto
 - área,
 - perímetro,
 - solidez,
 - excentricidad,
 - momentos centrales,
 - topología,
 - ...

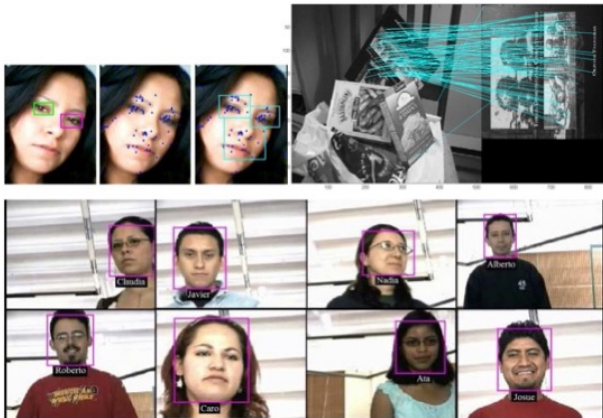
Ejemplo: caracterización de regiones para etiquetado automático



Características locales

- Dado que la segmentación es un problema muy complejo aún no completamente resuelto, una alternativa es evitar la segmentación
- Se utilizan ciertas características locales distintivas para el reconocimiento
- Se busca que dichas características sean robustas ante cambios de escala, rotación y traslación; así como ante cambios de iluminación
- Existen una gran variedad de características locales que se pueden utilizar:
 - Parches o templates
 - Orillas, esquinas (Harris)
 - Filtros sencillos (Haar)
 - Características invariantes (SIFT)
 - ...

Reconocimiento en base a SIFT



Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo
Visión de Nivel
Intermedio

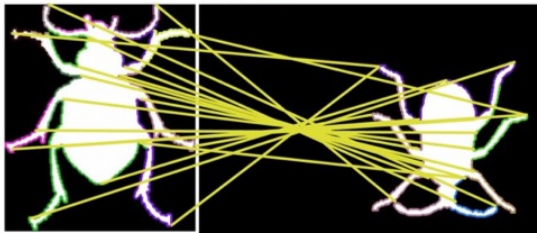
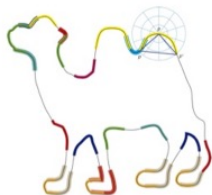
Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

Aplicaciones

Descriptores locales de forma



Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo
Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

Aplicaciones

Clasificadores

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo
Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

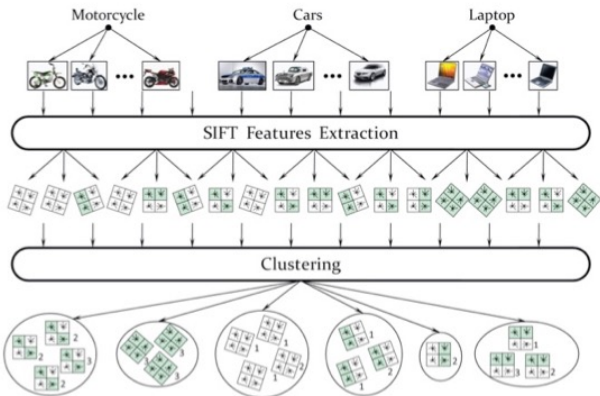
Aplicaciones

- Se busca maximizar la probabilidad del objeto o clase de objetos dadas las características:

$$P(C_i | X_1, X_2, \dots, X_n)$$

- Se pueden utilizar diversos tipos de clasificadores:
 - 1 Clasificador bayesiano simple
 - 2 Redes neuronales
 - 3 Árboles de decisión
 - 4 Ensamblajes de clasificadores
 - 5 ...

Ejemplo: reconocimiento en base a “palabras Visuales”



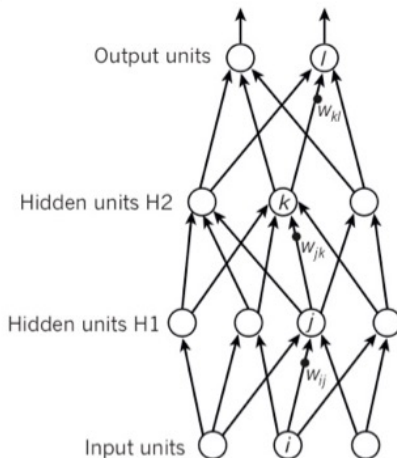
Visión Moderna

- En los últimos años ha cambiado el paradigma en visión computacional: en vez de utilizar técnicas para extraer características, segmentar, etc., se procesa directamente la imagen en *crudo*.
- Se basa en procesos de aprendizaje que aprovechan las grandes cantidades de datos (imágenes, videos) disponibles y la capacidades de cómputo, en particular las tarjetas gráficas o GPUs.
- Esto ha cambiado drásticamente el área; por ejemplo, en los principales congresos la mayor parte de los trabajos (95 %) utilizan aprendizaje profundo!

Aprendizaje Profundo

- El aprendizaje profundo o *deep learning* es una alternativa para la solución de diversos problemas en visión computacional
- Se basa en redes neuronales de muchas capas (5–30) que aprenden representaciones a varios niveles de las imágenes, usando usualmente un clasificador en la última capa.
- Requieren de una gran cantidad de datos de entrenamiento y buen poder de cómputo.

Red Neuronal Multicapa



Redes Neuronales Covolucionales (CNNs)

- Una arquitectura muy utilizada en visión.
- Se compone de una serie de capas alternativas de convolución y agrupamiento, que tienen cierta semejanza con el sistema visual biológico
- La convolución consiste de una serie de filtros que se aplican a toda la imagen, similares a los filtros clásicos, pero la diferencia es que sus coeficiente se aprenden
- El agrupamiento extrae los máximos de grupos de filtros

Red Neuronal Convolutacional

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

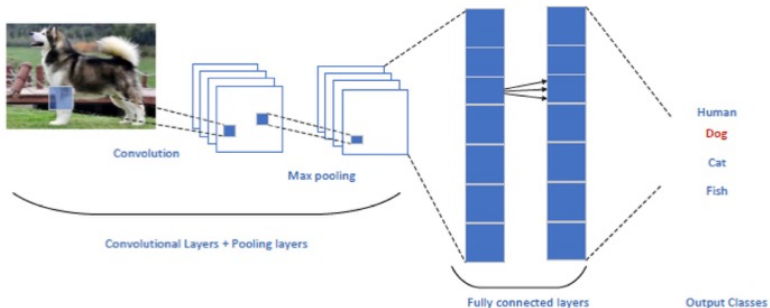
Visión Clásica

Visión de Nivel Bajo
Visión de Nivel Intermedio
Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

Aplicaciones



Arquitecturas

- Mediante el entrenamiento con grandes bases de datos se han generado diversas arquitecturas basadas en CNNs que sirven de base a diversas aplicaciones
- Estas arquitecturas o *backbones* se pueden adaptar a nuevos datos / aplicaciones modificando generalmente las últimas capas de la red neuronal, para lo que se requieren menos datos
- Varias se han publicado como librerías abiertas lo que ha ayudado al crecimiento del área

Arquitecturas y Aplicaciones

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

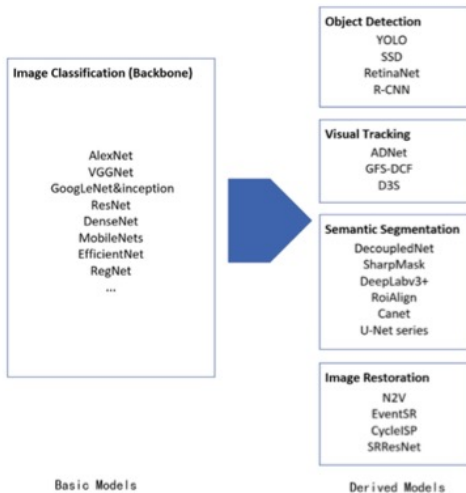
Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales

Transformers

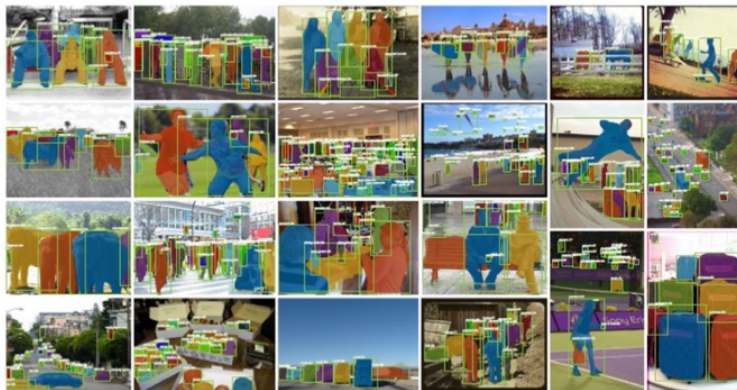
Aplicaciones



Aplicaciones de CNNs

- El uso de DL, en particular CNNs, ha tenido un gran impacto en mejorar los resultados en muchas aplicaciones.
- AlexNet en 2012 logró una mejora del 10.2 por ciento respecto al siguiente en el *ImageNet Large Scale Visual Recognition Challenge*, lo que desató el crecimiento de este enfoque.
- Se han logrado mejorar los resultados en varios problemas como clasificación, detección, segmentación, seguimiento, etc.

Segmentación



Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones

Seguimiento



Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo
Visión de Nivel Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

Aplicaciones

Transformers

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales

Transformers

Aplicaciones

- Los transformers son una nueva arquitectura de redes neuronales que aprenden relaciones de mayores rangos.
- Surgen inicialmente en el área de procesamiento de lenguaje, en particular en traducción automática.
- Recientemente se han empezado a utilizar también en visión, con buenos resultados en algunas aplicaciones.

Arquitectura

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel Intermedio

Visión de Alto Nivel

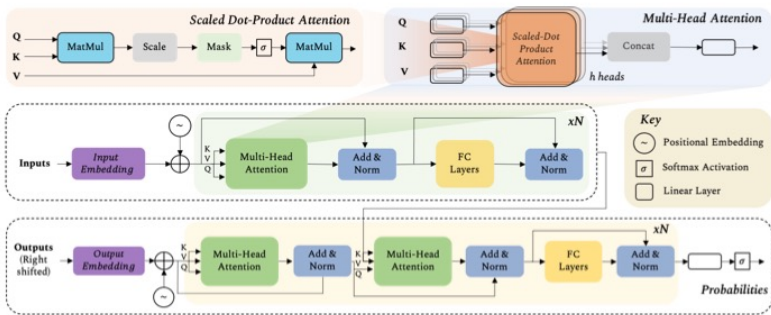
Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones



Transformers en Visión

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales

Transformers

Aplicaciones

- Para aplicar los transformers a visión, en vez de palabras, se utilizan regiones o *patches*.
- Esto permite incluir relaciones lejanas de la imagen, mientras las CNNs tienden a favorecer relaciones cercanas
- La idea central es este mecanismo de auto-atención (self-attention).
- En base a esta idea se han generado diversas variantes de la arquitectura de Transformers enfocados a visión.

Transformer para Visión

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

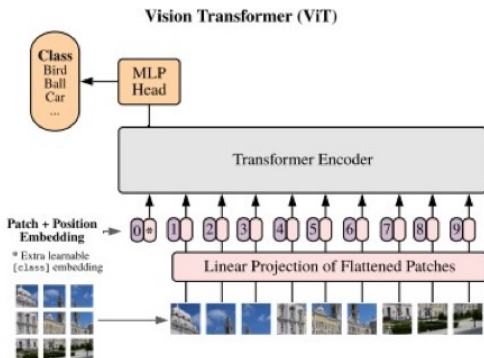
Visión Clásica

Visión de Nivel Bajo
Visión de Nivel
Intermedio
Visión de Alto Nivel

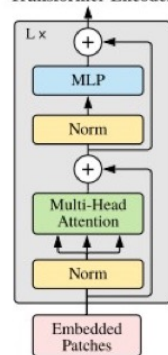
Visión
Moderna

Redes
Convolucionales
Transformers

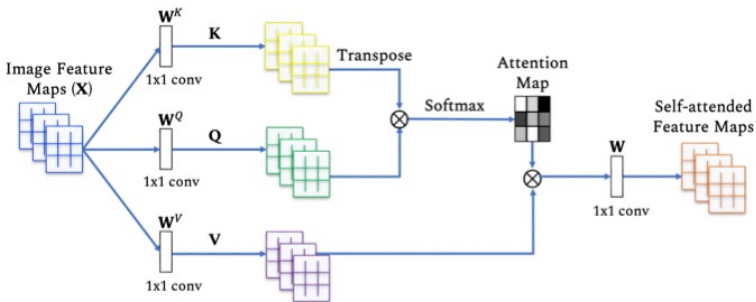
Aplicaciones



Transformer Encoder



Self Attention



Tareas en Visión

- Los transformers se han aplicado con éxito a diversas tareas: detección y reconocimiento de objetos, segmentación, super-resolución, entendimiento de videos, síntesis de texto–imágenes, etc.
- Una de las más impresionantes es la generación de imágenes a partir de frases de texto, DALL-E



Fig. 10: Images generated by DALL-E [20] from the following text prompts. (a) *An armchair in the shape of an avocado.* (b) *A photo of San Francisco's golden gate bridge. Given a part of the image (in green box), DALL-E performs the image completion.* (c) *An emoji of a baby penguin wearing a blue hat, red gloves, green shirt, and yellow pants.* (d) *An extreme close-up view of a capybara sitting in a field.* (e) *A cross-section view of a pomegranate.* (f) *A penguin made of watermelon.* (g) *The exact same cat on the top as a sketch on the bottom.*

Ejemplos de Aplicaciones

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales

Transformers

Aplicaciones

- Robótica
- Imágenes médicas
- Reconocimiento de personas
- Reconocimiento de actividades
- Video vigilancia
- Control de calidad
- Seguimiento (tracking)
- Anotación y recuperación de imágenes
- Reconocimiento de emociones

Robótica - reconstrucción 3D

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

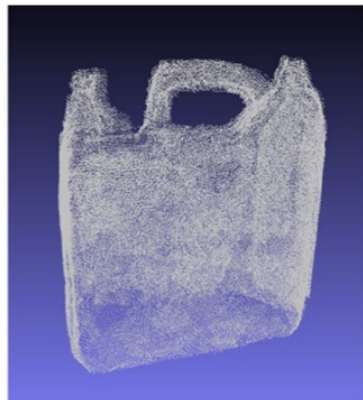
Visión de Alto Nivel

Visión
Moderna

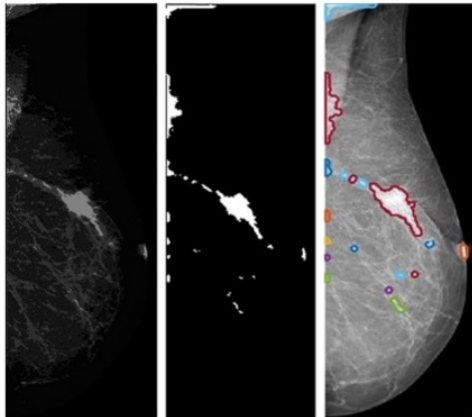
Redes
Convolucionales

Transformers

Aplicaciones



Imágenes médicas - segmentación de tumores en mamografías



Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones

Rehabilitación virtual

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

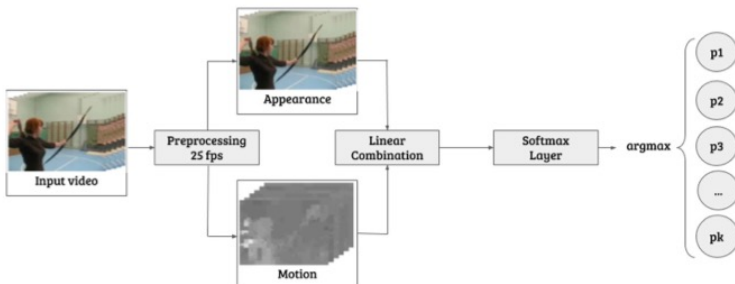
Redes
Convolucionales

Transformers

Aplicaciones



Reconocimiento de actividades en video



Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones

Reconocimiento de personas

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones



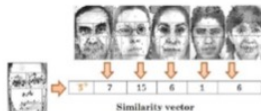
Localización y seguimiento



Resultados



Extracción de características

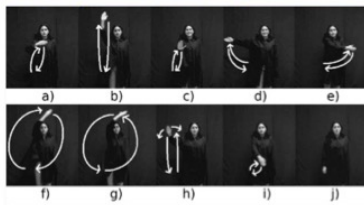
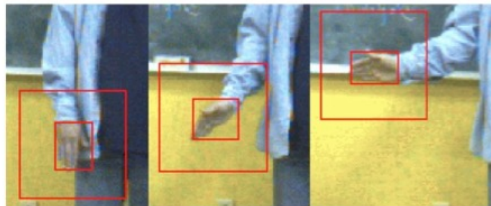


Reconocimiento

Acumulación de evidencia

$$P(f_i|s) = \frac{P(s|f_i)P(f_i)}{P(s)} = \frac{P(s|f_i)p(f_i)}{\sum_{k=1}^n P(s|f_k)P(f_k)}$$

Reconocimiento de Ademanos



Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales

Transformers

Aplicaciones

Detección de anomalías

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones



Reconocimiento de emociones

Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo
Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales
Transformers

Aplicaciones

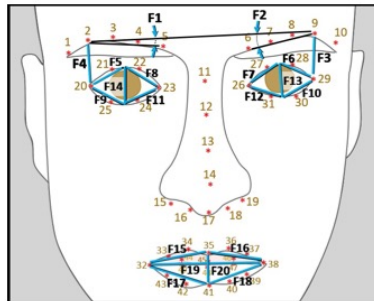


Anger Disgust Fear Joy

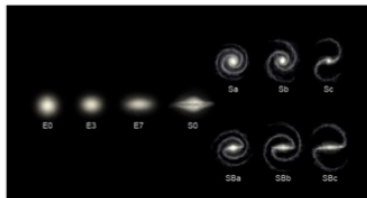


Neutral Sadness Surprise

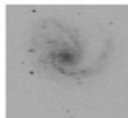
Fig. 1. Seven basic human emotions



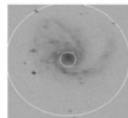
Clasificación de Galaxias



Nucleus separation



Original Galaxy



Galaxy with separated nucleus

Introducción a
Visión
Computacio-
nal

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales

Transformers

Aplicaciones

Video Vigilancia



Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión *Clásica*

Visión de Nivel Bajo

Visión de Nivel
Intermedio

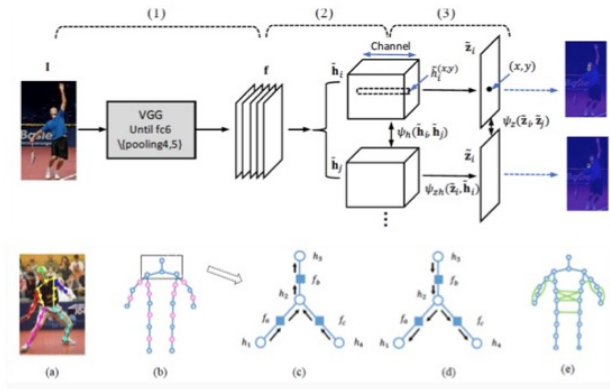
Visión de Alto Nivel

Visión
Moderna

Redes
Convolucionales
Transformers

Aplicaciones

Estimación de Postura - DL + RB



Introducción a
Visión
Computacional

Eduardo
Morales,
Enrique Sucar

Introducción

Visión Clásica

Visión de Nivel Bajo
Visión de Nivel
Intermedio

Visión de Alto Nivel

Visión

Moderna

Redes
Convolucionales
Transformers

Aplicaciones

Referencias (1)

- D. Marr, Vision, Freeman
- L.E. Sucar, G. Gómez, Visión Computacional (<http://ccc.inaoep.mx/esucar/Libros/vision-sucar-gomez.pdf>)
- Russel and Norvig, Cap. 24
- S. Ullman, High-Level Vision, MIT Press
- O. Faugeras, Three Dimensional Computer Vision, MIT Press

Referencias (2)

- Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Deep Learning, Nature 521, 2015
- Junyi Chai, Hao Zeng, Anming Li, Eric W.T. Ngai, Deep learning in computer vision: A critical review of emerging techniques and application scenarios, Machine Learning with Applications 6, 2021
- Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah, Transformers in Vision: A Survey, arXiv:2101.01169v5, Jan 2022