

Apuntes VAAC

Tema 1: Introducción

Luz y color: su percepción y representación

Espectro visible: región del espectro electromagnético que el ojo humano es capaz de percibir.

Luz visible: radiación electromagnética en este rango de longitudes de onda.

Un típico ojo humano responde a longitudes de onda de 400 a 700 nm.

Diagrama espectral: Informa de cómo está distribuida la potencia o la energía de una señal sobre las distintas frecuencias de las que está formada (distribución espectral).

Color: propiedad percibida que nos permite distinguir luces con diferentes longitudes de ondas (a distinta longitud se producen diferentes sensaciones de color). Se involucran 3 conceptos: *matiz*, *saturación* y *brillo*.

Brillo o iluminación:

- La luminancia es un término físico que permite medir la cantidad de luz y se mide en lúmenes
- El brillo y la iluminación son equivalentes en percepción a la luminancia.
- Se suele usar el término de iluminación para la luz reflejada por objetos.
- Se usa el término brillo para la luz emitida por fuentes luminosas.
- Cuando no hay matiz ni saturación se habla de percepción acromática.

Matiz: la frecuencia dominante en el diagrama espectral de una luz define su matiz.

Saturación: la diferencia entre el valor de la frecuencia dominante y el resto de frecuencias

Ojo: la retina es un tejido del interior del ojo que contiene células fotosensibles sobre la que se proyectan las imágenes exteriores:

- Los bastones son muy numerosos (100 millones) y los conos son menos numerosos (6 millones)
- Los bastones obtienen máxima respuesta a 507 nm y solo están activos en condiciones de baja iluminación por lo que se creen responsables de la visión nocturna.
- Conos: hay tres tipos de conos, cada uno sensible a una diferente longitud de onda (corta, medias y larga), son las células involucradas en la percepción cromática.

Percepción cromática (Teoría triestímulo): establece que mezclando luz de 3 linternas ideales (una roja, otra verde y otra azul) se obtienen todos los colores que somos capaces de percibir

RGB: el modelo RGB (por Red, Green, Blue) es un modelo aditivo de color en el que se

mezclan luces rojas, verdes y azules. Los ordenadores representan los colores mediante tripletas de valores RGB (3 bytes, uno para cada color, 16 millones de posibles colores).

HSV (o HSB): Inspirado en el círculo cromático es un sistema cilíndrico para la representación de colores.

Captura de imágenes: óptica y cámaras

Modelo ideal de cámara (pinhole)

- Una superficie mate refleja luz en todas direcciones.
- Con apertura muy pequeña, sólo hay un rayo de luz desde cualquier punto de la escena al plano imagen.
- Todos los puntos están bien definidos: imagen enfocada
- Si se aumenta la apertura, los puntos de la escena se difuminan en el círculo de confusión.
- Si el diámetro del círculo de confusión es menor que el tamaño de cada celda del fotodetector la imagen estará enfocada.

Modelo de lente fina

Lente: posee un plano de simetría, un eje de simetría denominado eje óptico, y un punto interior (centro óptico) donde se corta el eje óptico y el plano de simetría.

Foco: todos los haces paralelos que inciden perpendiculares a su plano de simetría, se cortan en un punto llamado foco.

Rayos por centro óptico: todo haz de luz que pasa por su centro óptico continúa en línea recta.

Plano de formación de la imagen: sea un punto P, que se encuentra a una distancia de la lente fina mucho mayor a la distancia focal. Todos los rayos que provengan de P, tras atravesar la lente, se cortan en un punto llamado punto de formación de la imagen.

Parámetros del modelo de lente fina: la lente permite más de un rayo de luz proveniente del mismo punto.

Profundidad de campo: rango de distancias a la cámara en la que los objetos aparecen enfocados.

Cámaras reales

Parámetros de cámaras:

- **Sistema de lentes (óptica utilizada):**
 - Aumento (zoom): establece el ángulo de visión y la distancia focal.
 - Enfoque: su ajuste, junto con la apertura, determina la profundidad de campo.
- Tamaño de apertura (o diafragma).
- Tiempo de exposición (o velocidad de obturación).
- Sensibilidad en los fotodetectores (o valor ISO).

Los parámetros se ajustan para conseguir 2 objetivos:

- Que entre la cantidad de luz suficiente.
- Que los objetos de interés estén enfocados.

Apertura del diafragma: cuanto menor es la apertura, más profundidad de campo pero entra menos luz.

Velocidad del obturador: tiempo durante el cual se deja pasar la luz al fotodetector (en segundos). Junto con la apertura, determina la cantidad de luz que entra.

Sensibilidad (valor ISO): desde 3200 ISO (muy sensible) hasta 50 ISO (poco sensible), cuanto más sensible, más ruido (efecto de granularidad).

Aberraciones de las lentes

Captura de imágenes digitales

Fotodetectores digitales

Tipos:

- CCD
- CMOS

Patrón de Bayer: distribución típica de los filtros de color en el array de fotodetectores (más detectores para el verde).

Limitaciones:

- Campo de estrellas. mayor cuanto peor es la cámara.
- Ruido fotónico
- Rebosamiento (blooming) solo CCD

CCD vs CMOS:

- CCD:
 - La carga de cada celda se transporta, a su vecino para convertirla en voltaje y después almacenarla.
 - Casi toda la celda se puede dedicar a la captura de luz.
 - La salida es muy uniforme, sobre todo con baja iluminación.
- CMOS:
 - Cada fotodetector tiene su conversión carga-voltaje, amplificador, corrección del ruido y a veces circuitos de digitalización.
 - Hay menos área para la captación de luz en cada celda.
 - La salida de una retina CMOS no es muy uniforme.
 - La salida es masivamente paralela, con gran ancho de banda y posibilidad de entregar sub-imágenes fácilmente.

Imagen digital

Es una representación matricial de una imagen como un conjunto finito de elementos (píxeles o "picture elements") que toman valor discreto.

La digitalización implica que la imagen digital es una aproximación a la imagen real.

Muestreo/resolución espacial: la digitalización comienza tomando muestras sobre el área del fotodetector de la cámara (conjunto discreto de celdas en una matriz).

Cuantificación: en cada celda del fotodetector se discretiza el voltaje correspondiente a una cantidad de luz y se asigna un valor binario a esa cantidad. El número de niveles (resolución digital) en los que se divida el voltaje máximo determinará el número de bits por cada píxel.

Definición de Imagen digital (grises y color): una imagen digital es una función

$I : F \times C \longrightarrow D$ Siendo $D = \{0, \dots, d - 1\}$ la resolución digital de I (típicamente $d=256$).

Una imagen digital en color es una función $I : F \times C \times P \longrightarrow D$ siendo $P = \{0, \dots, p\}$ el conjunto de planos de color.

Imágenes con más de 3 planos:

- El canal de transparencia (alpha o A), sirve para colocar una máscara que definirá que partes son transparentes o no.
- Imágenes multiespectrales: diferentes filtros a diferentes longitudes de onda de la luz que permiten descubrir diferentes materiales en la imagen.

Ejemplo de carga de imagen: `img1 = cv2.imread("Lichtenstein.png",1)`

Tipos de píxel (imagen):

- Imagen binaria: 1 píxel (1 bit, 0 negro 1 blanco)
- Imagen en escala de grises: 1 píxel (1 byte, 0 negro 255 blanco $0 < x < 255$ grises)
- Imagen en color: 1 píxel (3 bytes, 1 byte por color).

Compresión de imágenes: tres tipos de redundancia:

- En la codificación.
- En la representación espacial de los píxeles
- Visual.

Redundancia en la codificación: en una imagen suelen existir niveles de intensidad que son más probables que otros porque aparecen más veces. La codificación de tamaño variable es una técnica de compresión que aprovecha esta circunstancia asignando un código más corto a los niveles de intensidad más probables. (Un método de este tipo es la compresión Huffman)

Redundancia en la representación: se aprovechan de que se desea comprimir imágenes y se basan en las características propias de ellas. El método CCITT 3, aprovecha que es más redundante almacenar cada bit que rachas de blancos y negros. Las pirámides de Laplace ofrecen otro mecanismo de compresión que se conocen como sistemas predictivos.

Redundancia visual: cuantos menos coeficientes almacena, menos se parece a la imagen original, pero menos datos se almacenan.

Tema 2: Operaciones sobre imágenes

Introducción

Procesamiento de imágenes en VC: el procesamiento de imágenes transforma una imagen de entrada en otra imagen.

Operaciones sobre la imagen:

- Operaciones globales (se tiene en cuenta toda la imagen).
- Operaciones locales (se tienen en cuenta una vecindad del píxel)
- Operaciones puntuales (se tiene en cuenta su antiguo valor)

Operaciones puntuales

Artimético-lógicas con un plano

Artimético-lógicas con varios planos

Resta de imágenes: detección de movimiento en vídeo.

Media de imágenes: eliminar (o atenuar) el ruido gaussiano añadido a la imagen.

Histograma

Ejemplo OpenCV: `hist_item = cv2.calcHist([image],[0],None,[256],[0,255])`

Operaciones puntuales

Operaciones puntuales: las operaciones sobre imágenes más sencillas sólo tienen en cuenta el nivel de gris del píxel.

Imagen negativa: Realzar regiones blancas o grises en zonas muy oscuras (d-1 es el nivel de gris máximo).

Umbralización: Para separar (segmentar) zonas interesantes del fondo.

Asunción general: Al margen de la transformación de “negativo” y el “umbralizado”, en la mayoría de las operaciones puntuales, como se desea mantener la relación de orden entre los valores de los píxeles, se asume que: *T es una función monótona no decreciente*

Identidad/Negativo: histograma invertido.

Potencia y raíz cuadrada:

- Parábola: oscurecer los medios tonos.
- Raíz: aclarar los medios tonos.
- Dos trozos de curva (parábola y raíz): aclarar tonos oscuros y oscurecer los claros.

Corección gamma: el humano no responde de manera lineal a los cambios lineales de luminancia. El cambio perceptivo al pasar de contemplar 1 bombilla encendida a 2

bombillas es mucho mayor que el percibido al pasar de 10 bombillas a 11. Por eso, en los sistemas informáticos actuales las imágenes se almacenan codificadas con una gamma de 0.45 y se decodifican con una gamma de 2.2.

Rango dinámico: conjunto de valores de intensidad presentes en la imagen.

Contraste: Diferencia en intensidades entre el máximo valor de intensidad presente en la imagen y el mínimo (a mayor rango dinámico, la imagen tiene más contraste).

Aumento del contraste: encontrar una transformación que aumente el rango de niveles de gris, y por tanto el contraste, de la imagen:

- Transformación lineal.

Operaciones aritméticas

Se puede:

- Sumar/restar una constante
- Multiplicar/dividir por una constante

Ecualización

Procedimiento automático de cálculo de $T[r]$ de tal forma que en la imagen transformada:

- El rango dinámico sea el máximo
- Los niveles de gris se distribuyan uniformemente

El objetivo del ecualizado es transformar la distribución del histograma P_r en una distribución uniforme P_s .

Cualquier transformación del histograma cumple que el área a la izquierda de un punto r sobre la distribución original P_r es igual al área a la izquierda del correspondiente punto transformado s en la distribución P_s (el área (a) bajo P_r es igual a (a) bajo P_s).

Ejemplo de ecualización en OpenCV: `hist_item = cv2.calcHist([image],[0],None,[256],[0,255])`
`pyplot.plot(hist_item.reshape(256,))`

Operaciones de vecindad

Se aplica a los píxeles alrededor del píxel actual.

Tipos de operaciones de vecindad:

- Lineales (filtrado espacial).
- No lineales.

Filtrados espaciales

Filtrado en el dominio del espacio: T es una combinación lineal de los píxeles vecinos al (i,j) (operador de correlación entre f y la máscara p).

Operador de convolución: la convolución entre f y W se denota como $w \cdot f$ y es igual que

la correlación pero invirtiendo la máscara.

Detalles de implementación: el rango: para evitar que el resultado de una operación de convolución sobre una imagen se salga de rango se le suelen aplicar dos correcciones:

- Factor divisor: que divide al resultado por un valor.
- Factor bias: que se le suma un valor al resultado.

Detalles de implementación: los bordes:

- La máscara de filtrado se sale de la imagen y es necesario extrapolar y hay diferentes soluciones:
 - Dejar el borde negro (sin hacer nada más)
 - Considerar que el siguiente píxel al 0 es el n-1 y viceversa.
 - Copiar el borde tantas veces como se necesite.
 - Reflejar la imagen en la dirección de cada borde

Propiedades de la convolución:

- Es lineal
- Conmutativa
- Asocitativa
- Identidad

Filtros separables: sea h una máscara de convolución (matriz) de dimensión $m \times n$, la condición necesaria y suficiente para que h sea separable es que $\text{rango}(h)=1$.

Suavizado (filtrado paso bajo)

- Dejan pasar las frecuencias bajas (zonas de sombreado uniforme).
- Eliminan (en parte) las frecuencias altas (zonas donde hay cambios rápidos en los niveles de gris).
- El ruido suele ser de frecuencia alta (¡y los bordes también!)

Filtrado espacial en OpenCV: `filtered_image = cv2.filter2D(image,-1,kernel) # Blur`
hace lo mismo

Filtro de promediado:

$$\frac{1}{9} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \frac{1}{3} \cdot \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \frac{1}{3}$$

Promediado
("box filter")

La máscara de promediado es separable

Filtro gaussiano:

$$\frac{1}{16} \cdot \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} = \frac{1}{4} \cdot \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} * \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \cdot \frac{1}{4}$$

Aproximación a la gaussiana
("gaussian filter")

La máscara gaussiana es separable

Filtro de promediado y gaussiano: El filtro de promediado no representa bien el proceso de desenfoque de una lente. Un punto blanco se transforma en un cuadrado con lados desenfocados (no en un punto desenfocado).

Filtro gaussiano: Es mejor utilizar una máscara $m \times n$ que dé más importancia a los píxeles más cercanos al píxel central. La distribución Gaussiana (normal) de media $(0,0)$ en 2D hace eso mismo. El filtro gaussiano es separable.

Varianza del filtro gaussiano: determina la fuerza del suavizado.

Tamaño de la máscara gaussiana: En una gaussiana el 99,7% de la densidad probabilidad está entre -3σ y 3σ . Por tanto el tamaño del filtro espacial n suele tomarse $\geq 6\sigma$.

Filtrado gaussiano en OpenCV: `filtered_image = cv2.GaussianBlur(image,(31,31),5)`

Ruido en los niveles de gris:

- Ruido de "sal y pimienta": muestras aleatorias de píxeles negros y blancos.
- Ruido impulsivo: ocurrencias aleatorias de píxeles blancos.
- Ruido gaussiano: variaciones en el nivel de gris debidos a una distribución gaussiana.

Detectar bordes (filtrado paso alto): el objetivo es identificar discontinuidades en la imagen (cambios bruscos en el nivel de gris). En los bordes está la mayor parte de la información sobre los objetos. Es una información más elaborada (y compacta) que los

píxeles.

Suavizado vs bordes:

- Suavizado, es hacer la media
- Buscar bordes, es enfatizar las zonas de la imagen con cambios fuertes en el nivel de gris

Bordes y la derivada del nivel de gris: un borde es una zona de la imagen con cambio rápido en el nivel de gris de la imagen.

Algoritmo de Canny (1986): se usa para calcular bordes con la derivada primera, norma de vector de gradiente, etc.

Operador Laplaciana (Laplacian): El operador de la Laplaciana combina las dos derivadas segundas. Es isotrópico (misma respuesta independientemente de la orientación del borde). La suma de la derivada primera en dirección X y en dirección Y para un punto podría cancelarse si tuviesen signos opuestos. Como esto no ocurre para la derivada segunda, podemos agrupar ambos términos en la misma ecuación.

Los bordes y la escala (I): Cuando suavizamos con una gaussiana antes de calcular la derivada, ¿qué valor de deberíamos utilizar? La desviación típica de la gaussiana define la escala de los bordes que vamos a poder detectar. Con un pequeño σ detectaremos todos los bordes, con un grande σ sólo detectaremos el borde mayor.

Operaciones morfológicas

Morfología matemática: se basa en operaciones de teoría de conjuntos, la imagen se ve como un conjunto de puntos. De momento solo se considera el caso de las imágenes bitonales, que se expresarían como un conjunto de pares de coordenadas correspondientes a los píxeles activos.

Definiciones:

Traslación de A por $X = (x_1, x_2)$, como:

$$(A)_x = \{c / c = a + x, a \in A\}$$

Reflexión de A como:

$$\hat{A} = \{x / x = -a, a \in A\}$$

Complementario de A como:

$$A^c = \{x / x \notin A\}$$

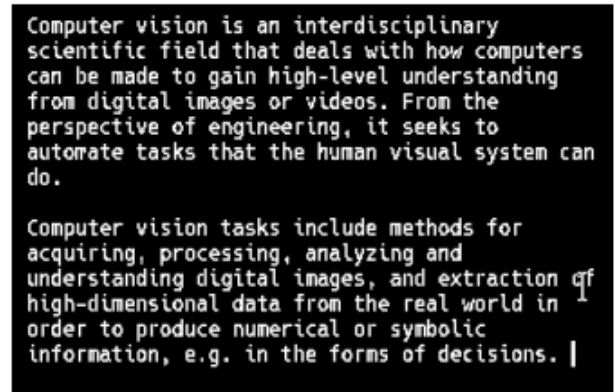
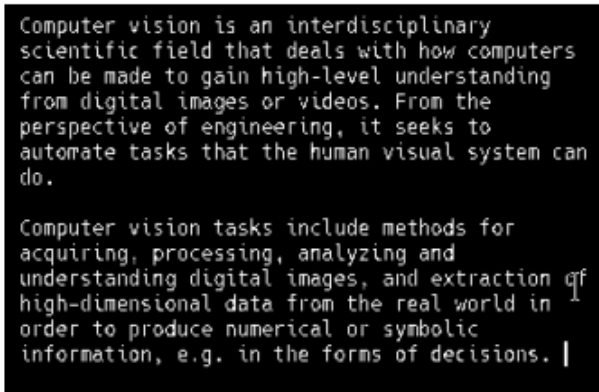
También se define la operación **diferencia** entre dos conjuntos A y B como:

$$A - B = \{x / x \in A \text{ y } x \notin B\}$$

Una propiedad interesante que se deriva de las operaciones anteriores es:

$$A - B = A \cap B^c$$

Idea intuitiva de la dilatación: Intuitivamente, para todo píxel activo de A se sitúa el centro de E sobre él y se suman a A todos los píxeles activos de E.



Idea intuitiva de la erosión: Tras una dilatación, interesa disponer de una operación contraria, que las deshaga, dentro de lo posible. Intuitivamente, para todo píxel activo p de A se sitúa el centro de E sobre él y si todos los píxeles activos de E no coinciden con píxeles activos de A, el píxel p se desactiva.

Operaciones geométricas

Las transformaciones geométricas cambian el dominio (las coordenadas) de la imagen.

Traslación 2D en forma matricial: las traslaciones en 2D suponen añadir una cantidad a cada coordenada del píxel.

Transformación rígida euclídea: consiste en una rotación, en sentido antihorario, seguida de una traslación.

Transformación de similaridad: en este caso la transformación es una rotación seguida de un escalado y una traslación.

Transformación afín: tenemos 6 incógnitas, con lo que con 3 puntos en f (imagen origen) y 3 puntos en g (imagen destino) tenemos 6 ecuaciones.

Transformación proyectiva (homografía): en este caso tenemos 8 incógnitas (grados de libertad) con lo que necesitaremos 4 puntos en f y 4 en g para estimar H . Por cada par (i, j) , (i', j') tenemos dos ecuaciones.

Opciones de escalado

Pirámide a diferentes escalas: En muchas ocasiones interesa reducir el tamaño de la imagen. Lo usual en visión es reducir a la mitad en cada dimensión. Ejemplo: cuando queremos buscar caras, peatones, etc. a diferentes escalas es habitual crear una pirámide de imágenes a diferente resolución.

Reducción de tamaño:

- submuestreo: Idea 1: eliminar 1 de cada dos columnas y 1 de cada 2 filas.

- Filtrado gaussiano previo: Idea 2 (filtrado + submuestreo): filtrado gaussiano previo y después eliminar 1 de cada dos columnas y 1 de cada 2 filas. El tamaño del filtro se dobla cada vez que hacemos una reducción de tamaño a 1/2

Aliasing: El aliasing es el efecto que causa que señales continuas distintas se tornen indistinguibles cuando se muestrean digitalmente. ¿Qué hacer contra el aliasing?:

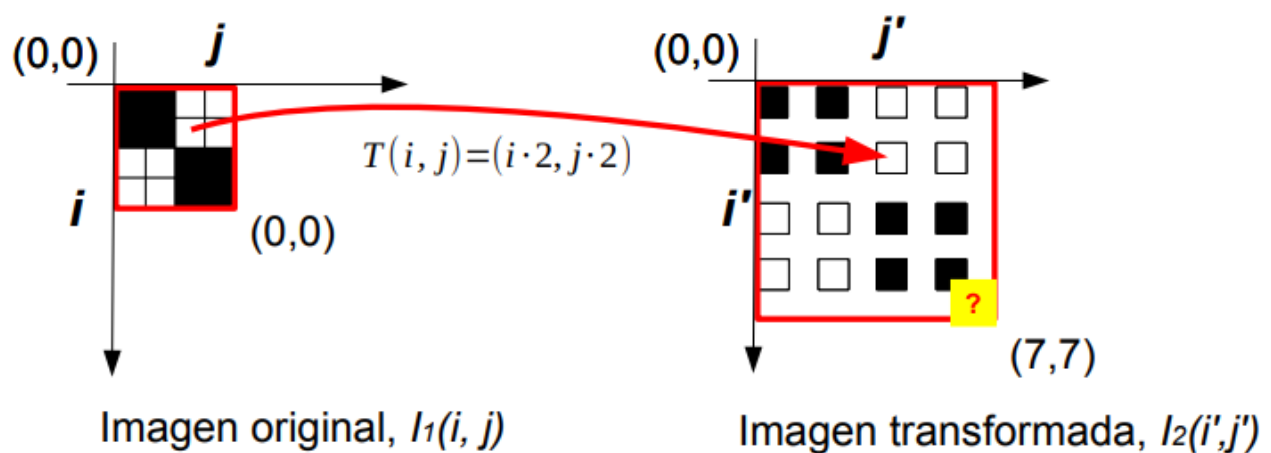
- Muestrear más a menudo (a mayor frecuencia)
- Hacer la señal menos “oscilante”

Teorema del muestreo: cuando se muestrea una señal a intervalos discretos, la frecuencia de muestreo debe de ser mayor que 2 veces la mayor frecuencia de la señal de entrada para poder reconstruir perfectamente la señal original a partir de la muestreada.

Aumento de tamaño: También es interesante incrementar el tamaño de una imagen. En este caso ya hemos perdido la información de alta frecuencia y queremos “recuperarla” ¿será eso posible?

Función de transformación directa: T

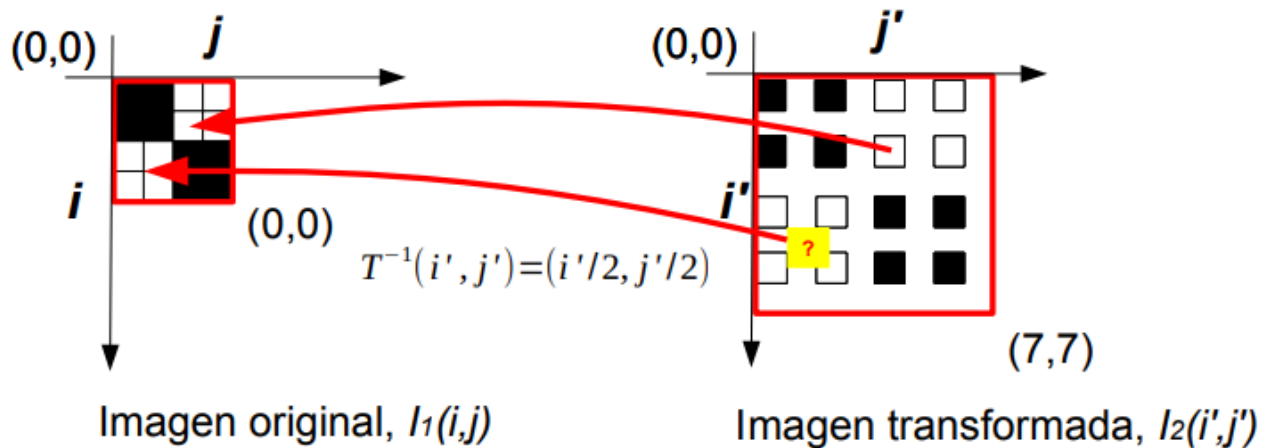
- Si utilizamos la transformación directa **¡quedan huecos en la imagen transformada!**



**¿Qué hacemos con los
píxeles para los
que no tenemos valor?**

Función de transformación inversa: T^{-1}

- Podemos recorrer la imagen transformada y encontrar el nivel de gris en la imagen original **por interpolación**.



¿Qué hacemos con los
píxeles para los
que no tenemos valor?

Tipos de interpolación:

- al vecino más próximo
- bilineal
- bicúbica

Tema 3: Detección y descripción de puntos de interés

Introducción a los puntos de interés

En muchas aplicaciones de visión es importante encontrar las correspondencias de ciertos elementos característicos en dos o más imágenes.

Aproximaciones globales y locales:

- aproximación local: tratará de encontrar los elementos característicos del objeto.
- aproximación global (holística): tratará de emplear una imagen del objeto completo.

Ventajas de la aproximación local:

- Restringidas a una pequeña región: Al estar localizadas, son robustas a oclusiones y fondo complicado.
- Distinguibles: pueden distinguir una gran cantidad de objetos.

- Cantidad: cientos o miles en una sola imagen .
- Eficiencia: se pueden obtener en tiempo real (15-30 imágenes/segundo)

Aplicaciones de puntos de interés:

- Alineación de imágenes (p.ej. creación de imágenes panorámicas)
- Reconstrucción 3D
- Seguimiento de objetos en movimiento.
- Navegación de Robots (SLAM visual).
- Indexado y recuperación en bases de datos de imágenes.
- Reconocimiento de objetos y rasgos biométricos.

Detección de puntos de interés

Objetivo: características locales que sean invariantes a transformaciones:

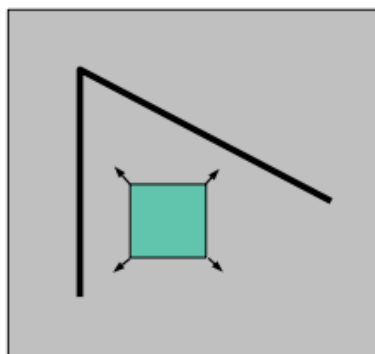
- Geométricas: traslación, rotación, escala.
- Iluminación: intensidad, dirección, etc

¿Qué puntos que podemos detectar?: buscamos regiones que sean distinguibles.

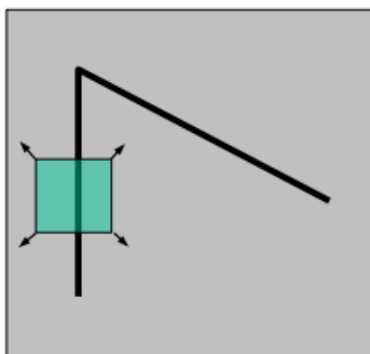
Detector de Harris: Medida local de “ser distinguible”: ¿Cómo cambian los niveles de gris de la ventana cuando esta se desplaza un poco?

Detector de Harris

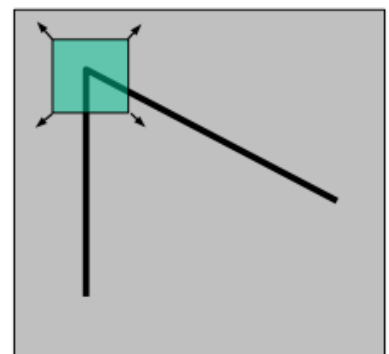
$E(u,v)$ = “grado de cambio al trasladar (u,v) la ventana”



$E(u,v)$ es pequeña
para cualquier
desplazamiento



$E(u,v)$ es pequeña
para algunos
desplazamientos



$E(u,v)$ es siempre
grande para cualquier
desplazamiento.

Propiedad clave: alrededor de una esquina el gradiente de los niveles de gris tiene dos direcciones dominantes.

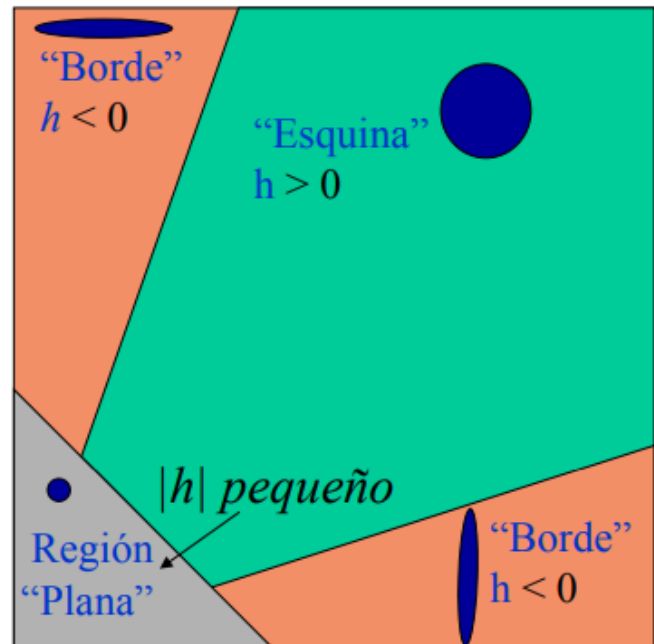
Valor de “esquinosidad” de Harris:

Valor de “esquinosidad” de Harris

$$h = \lambda_1 \cdot \lambda_2 - k \cdot (\lambda_1 + \lambda_2)^2 = \det(M) - k \cdot \text{traza}(M)^2$$

k : constante que se determina empíricamente y toma valores entre (0,04 y 0,06)

El cómputo exacto de los valores propios es costoso, Afortunadamente no es necesario computar la descomposición en valores propios de la matriz M y en cambio es suficiente evaluar el determinante y la traza de M



Detec. de Harris: Invarianzas:

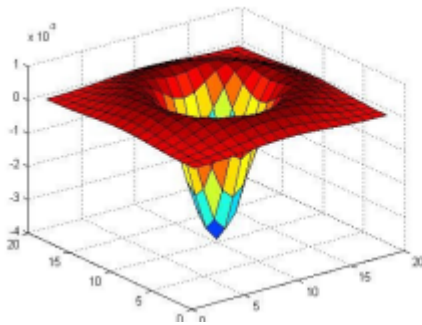
- La medida de Harris es invariante a la rotación de la imagen.
- La medida de Harris es parcialmente invariante a los cambios lineales en la intensidad (por el umbral).
- Harris no es invariante al cambio de escala de la imagen

Detección de P.I. invariante a escala: Objetivo:

- Encontrar la escala “característica” de la región de interés en imágenes tomadas a diferente escala.
- Diseñaremos una función que dependa de la región que estemos mirando y que sea “invariante a escala” (misma respuesta para regiones que se correspondan aunque estén tomadas a diferentes escalas).

El operador Laplaciana (otra vez): localización en la imagen: la magnitud de la respuesta de la Laplaciana es mínima/máxima en el centro de un “bulto”, si la escala de la Laplaciana “encaja” con la escala del “bulto”

La Laplaciana normalizada: La Laplaciana de una Gaussiana, con desviación σ , normalizada frente a escala tiene la siguiente forma:



$$\text{LoG}_{\text{norm}}(\sigma) = \sigma^2 \left(\frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} \right)$$

Harris-Laplace: encontrar las esquinas de Harris multiescala utilizando una matriz M modificada para tener en cuenta la escala. ¿Por qué no usar el valor de “esquinoidad”, H , del HarrisLaplace para encontrar un máximo local a lo largo de los diferentes valores de escala? ¿Por qué se utiliza la Laplaciana normalizada? Experimentalmente, Mikolajczyk y Schmid, descubrieron que la “esquinoidad” no generaba máximos locales en el espacio de escalas posibles.

Eficiencia en la Laplaciana: DoG: Aproximar la Laplaciana con una diferencia de Gaussianas es más eficiente que convolucionar con la máscara LoG, ya que la máscara gaussiana es separable.

Detectores de P.I. en OpenCV:

- SIFT (Scale Invariant Feature Transform): detector de esquinas que abrió camino para los detectores locales invariantes a escala y orientación
- SURF (Speeded Up Features): es un detector que es similar a SIFT pero que aproxima las derivadas de la gaussiana para que el cálculo sea mucho más eficiente.
- FAST (detector de esquinas): es un detector de esquinas pensado para los dispositivos de muy baja capacidad de cómputo (p.ej. Smartphones). Está basado en comparaciones rápidas del nivel de gris de 16 píxeles alrededor del centro de la esquina candidata.
- MSER (Maximally Stable Extremal Regions): detector de regiones de alto contraste a partir de probar diferentes umbrales en el nivel de gris.

Descripción de los puntos de interés

Descriptores de P.I.: vector de medidas que representan los niveles de gris en un entorno del Punto de Interés:

- Robustos a:
 - Transformaciones afines
 - Iluminación
 - Ruido
- Distinguibles (que identifiquen al P.I. frente a otros muy diferentes).
- Fáciles de emparejar:
 - Dimensión no demasiado grande.
 - Usualmente emparejados por distancia L1 (taxista) o L2 (euclídea).

Asignación de la orientación: Hasta ahora tenemos un conjunto de buenos Puntos de Interés. Ahora buscamos eliminar los efectos del cambio de orientación y escala en las imágenes.

Asignación de la orientación en SIFT:

1. Crear un histograma de ángulos de gradientes (36 entradas).
2. El pico máximo del histograma de gradientes nos dará la dirección u orientación dominante (perpendicular al borde dominante).

Solución SIFT: Orientación de los gradientes (independientes de la iluminación) en un histograma (evita los pequeños errores en i , j y θ)

Descriptor SIFT:

- Rejilla con 4x4 celdas, colocada sobre el P.I. (i,j), con la orientación q .
- En cada celda tomar 4x4 muestras (orientaciones de gradientes) y construir un histograma con 8 posibles orientaciones.
- Para calcular los histogramas se pesa cada muestra por una gaussiana alrededor de su centro (σ es 0,5 veces la escala del P.I.)
- 4x4 celdas con un histograma de 8 posiciones: $4 \times 4 \times 8 = 128$ dimensiones del descriptor.

Emparejamiento de descriptores SIFT:

- Distancia euclídea (L2) entre descriptores (vectores de 128 dimensiones).
- Posibilidades para emparejar descriptores (Puntos de Interés):
 - Umbral en la distancia entre ellos – Mal comportamiento.
 - Vecino más cercano entre muchos candidatos – Mejor
 - Dist. al más cercano entre dist. al segundo más cercano – El mejor

Descriptor BRIEF (binario, Binary Robust Independent Elementary Features):

A favor:

- Descriptor compacto (pocos bits), fácil de calcular, muy discriminante.
- Rápido de emparejar utilizando la distancia de Hamming.
- Buena tasa de reconocimiento.

En contra:

- Muy sensible a las distorsiones y transformaciones de la imagen (en particular a escala y orientación).

Está compuesto de pruebas de resultado binario en los niveles de gris. Y para cada P.I. que abarca una región de $S \times S$ píxeles:

1. Suavizar la región (con una máscara gaussiana 9×9 , $\sigma=2$)

2. Elegir los test binarios siguiendo un patrón pre-establecido.

Oriented Fast and Rotated BRIEF (ORB): Algoritmo propuesto por PenCVLabs (SIFT y SURF son algoritmos patentados, mientras que ORB no). Consiste en:

- Uso del detector de puntos FAST
- Uso del algoritmo de medida de Harris para obtener los N puntos con valor de medida más altos
- Uso de una pirámide de imágenes para producir características multiescala.
- Para obtener la orientación calcula un centroide ponderando la intensidad de gris de los puntos de FAST.
- El vector que va del punto detectado por FAST a dicho centroide.
- Uso de descriptor de puntos BRIEF rotado respecto a la orientación detectada.

Detectores de P.I: Conclusiones:

- Hay detectores/descriptores independientes de la escala y la orientación (hasta cierto punto): SIFT o SURF.
- Hay detectores de P.I. como Harris (o FAST) que no son independientes de la escala (aunque sí de la orientación).
- La independencia de la escala y orientación viene a cambio de menor velocidad en el cálculo.
- Los descriptores “pesados” como los de SIFT (o SURF) consisten en un vector de números reales a comparar con L2 (dist. Euclídea).
- Los descriptores “ligeros” como BRIEF (BRISK y otros) consisten en un número entero de 1 o 2 bytes que se pueden comparar con distancia de Hamming (un XOR bit a bit y una suma de los bits resultantes).

Video Google: Bag of Words: representación de un documento mediante la frecuencia de aparición de las palabras de un diccionario. Las “Palabras visuales” son fragmentos de imágenes. Se representa la imagen utilizando la frecuencia de aparición de palabras visuales (no su localización en la imagen).

Bag of Words: Algoritmo:

1. Se extraen Puntos de Interés
2. Se aprende el vocabulario visual (se agrupan los descriptores para obtener el representante de cada grupo o “palabra visual”).)
3. Cuando entra una imagen nueva se calculan descriptores y se asigna cada una a la palabra visual más cercana.
4. Se representa una imagen por la frecuencia de cada palabra visual.

Video Google: “query”:

1. Se representa una imagen “query” por la frecuencia de cada palabra visual: vector Q.

2. Se buscan imágenes en la base de datos que contengan las palabras visuales de la “query” (diccionario).
3. Se calcula el ángulo que forma el vector Q con el vector de frecuencias de las mismas palabras en cada documento D_i

Tema 4 Biometría

Biometría (o “Bioestadística”): se aplica a la ciencia que se dedica al estudio estadístico de las características cuantitativas de los seres vivos. Dicho término se utiliza también para referirse a métodos que analizan determinadas características humanas para reconocer y/o verificar la identidad de las personas. Uso de rasgos físicos o de comportamiento para la identificación de individuos. Para diferenciarlos se usará el término **Biometría Informática**.

Modos de comprobación de identidad (en sistemas de seguridad):

- ERES LO QUE POSEES (Algo que una persona tiene). Problemas: pérdida, robo, duplicación, puede compartirse, ...
- ERES LO QUE CONOCES (Algo que una persona sabe). Problemas: olvido, puede compartirse, puede adivinarse, ...
- ERES LO QUE ERES (Algo que una persona es). Ventajas: no puede compartirse, no puede perderse, ...

El puzzle biométrico (Jain, 2013): Asociar la identidad de personas con sus correspondientes rasgos biométricos (físicos y/o conductuales), y usar dichos rasgos en la verificación o el reconocimiento de personas.

Tratamiento de los datos ha de ser proporcional al fin que se busca. Privacidad de la identidad:

- El almacenamiento de datos biométricos es uno de los riesgos más graves de pérdida de privacidad
- La unión entre los datos de identidad biométrica y otros datos puede permitir vincular información sensible.
- Es crucial que los datos biométricos estén separados y protegidos de forma segura.

LOPD y GDPR

Rasgo biométrico o modalidad: Patrón único que identifique a cada individuo basándose en características físicas o de comportamiento:

- Modalidades fisiológicas (características físicas, ej huella, iris)
- Modalidades conductuales (características de actividad o comportamiento, ej voz, escritura)

Ejemplos:

- Aplicaciones forenses (huella dactilar, ADN,...)

- Video-vigilancia (imagen facial)
- Control de fronteras (huella dactilar, imagen facial, iris)
- Sistemas de control de acceso físico a recintos (huella dactilar, geometría de la mano, iris)
- Reconocimiento de clientes en aplicaciones bancarias (firmas)
- Autenticación remota (huella dactilar, imagen facial, voz)
- Control de acceso lógico a un ordenador (huella dactilar)

Características del rasgo biométrico:

- Universalidad : Todos los individuos deben poseer el rasgo.
- Singularidad o unicidad : Discriminación clara entre un individuo y el resto.
- Ponderabilidad : Debe poder medirse de forma cuantitativa
- Persistencia y estabilidad (permanencia) : Conservarse en el tiempo (Si tienen un periodo de estabilidad corto debe actualizarse su registro cada cierto tiempo (rasgos faciales))
- Rendimiento y calidad : Datos claros, precisos y analizables sin costosos preprocesos.
- Usabilidad : Fácil adquisición
- Vulnerabilidad : Resistente frente a ataques (suplantaciones u ocultaciones).
- Privacidad : Acceso restringido y si es necesario almacenarlo debe ser posible encriptarlo.
- Aceptación : Consentida por el individuo y no suponer un agravio
- Escalabilidad : Los procesos rápidos en sistemas con gran número de usuarios
- Mantenimiento : Sensores sin demasiados cuidados
- Integración : Compatibilizar la identificación biométrica con otro tipo de identificación.

La usabilidad no solo es la colocación de dedos o caras ante sensores biométricos. Abarca otros aspectos de la interacción persona-computador.

El contexto (lugar, tiempo y actividad) asociado a un sistema biométrico tiene gran influencia en su grado de aceptación.

Grado de aceptación diferente para una misma aplicación según países.

Imagen facial:

- Pros:
 - Conveniente
 - no intrusiva
 - barata
- Contras:
 - Baja precisión

Iris:

- Pros:

- Rendimiento muy alto
- Contras:
 - Intrusivo
 - No conveniente
 - Caro

Huellas dactilares:

- Pros:
 - Conveniente
 - no intrusiva
 - alto rendimiento
 - barata (sensores económicos)
- Contras:
 - No fiabilidad en todos los escenarios
 - Dificultades de autenticación para algunos usuarios
 - Facilidad de spoofing

Multimodalidad: Combinación de información de varios rasgos, motivada por las limitaciones de los sistemas unimodales:

- Ruido en la adquisición: sensores defectuosos o malas condiciones de adquisición.
- Variabilidad intra-usuario : interacción incorrecta con el sensor, variación de las características del sensor, o variación de las características del usuario.
- Menor poder discriminante: información limitada según cada rasgo
- No universalidad: fallo al registrar a determinados usuarios.
- Ataques al sistema: uso de rasgos artificiales.

Ventajas de la multimodalidad:

- Varias fuentes de información biométrica se combinan para mejorar el rendimiento de un determinado sistema.
- Se aumenta la cobertura de la población reduciendo el fallo en el registro.
- Se mejora la seguridad del sistema al aumentar la dificultad de imitar o falsificar varios rasgos simultáneamente.

Resumen del uso de la biometría:

- Ventajas:
 - No puede ser fácilmente transferida, olvidada, perdida o copiada
 - Alta precisión de algunas modalidades biométricas
 - Mayor protección ante el fraude
- Desventajas:

- El resultado depende de un umbral de decisión
- No puede ser “reseteada” ante situaciones problemáticas
- La seguridad de los sistemas biométricos pueden ser atacada
- Problemas de privacidad

Falsas afirmaciones sobre la biometría:

- La modalidad biométrica (o biometría) X es “la mejor” para todas las aplicaciones
- Una biometría X es única para todos los individuos.
- Un simple número cuantifica la precisión del sistema.
- La multimodalidad biométrica siempre mejora los resultados de usar una sola modalidad.
- Un mayor número de patrones biométricos (templates) implican una mayor precisión.
- Los sistemas biométricos ofrecen una seguridad absoluta.

“Desafío biométrico”: Encontrar una representación (extraída de la modalidad seleccionada) y una medida de similitud tal que

similitud **intra-personal** sea muy **alta**.
similitud **inter-personal** sea muy **baja**.

- **similitud intra-personal** : Capacidad identificar a un **individuo** a pesar de las **variaciones** en las características biométricas debidas a factores como la edad, la iluminación y la calidad de la imagen.
- **similitud inter-personal** : Capacidad para distinguir entre dos **individuos diferentes**.

Un sistema biométrico busca conseguir dos objetivos, bien la identificación o bien la verificación de la identidad de un sujeto.

Las etapas (módulos) de un sistema biométrico:

- inscripción o registro del sujeto (enrollment)
- identificación/verificación del sujeto

Existe un total de cuatro posibles situaciones del sistema:

- True positive
- False positive
- True negative
- False negative

Las tasas FAR (False Acceptance Rate) y FRR (False Rejection Rate) son funciones del grado de seguridad deseado (umbral de aceptación)

Curva ROC:

- El área bajo la curva (AUC) indica la “calidad” del sistema
- En las curvas ROC no se representa el valor del umbral.
- Se pueden representar varios algoritmos en la misma grafica con objeto de compararlos.

Reconocimiento facial

Imagen facial = principal rasgo (biométrico) usado por los seres humanos en la interacción inter-personal. Los sistemas biológicos emplean sensores visuales y mecanismos de interpretación de la información visual. De manera similar, los sistemas de visión artificial usan diferentes tipos de sensores y algoritmos para el reconocimiento facial.

Problemas asociados:

- Localización de la cara - variaciones de pose
- Variaciones de iluminación
- Expresiones faciales
- Oclusiones faciales
- Resolución y calidad de la captura
- Variaciones intra-clases frente a variaciones inter-clases, variaciones en el tiempo (p.ej. diferentes edades)

Variaciones de la pose: Una imagen 2D es una posible de las múltiples proyecciones, mucha variabilidad, etc. Uso de imágenes 3D para minimizar estas dificultades.

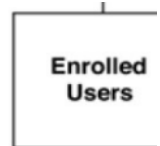
Cambios en la iluminación: Diferentes tipos y orientaciones de la iluminación de la escena producen variaciones en la apariencia facial. Solución: corregir parcialmente el efecto de la iluminación, aprender las variaciones de iluminación de los patrones, uso de imágenes 3D, etc.

Expresiones faciales: Las caras expresan emociones e intervienen múltiples músculos faciales cuyo movimiento determina las expresiones y hace variar la apariencia

Oclusiones faciales: En muchas aplicaciones, la oclusión parcial de caras es común,

Causada por el vestido (p.ej. bufanda), uso de gafas, peinado, autooclusiones.

Etapas principales:



- **Captura** (o adquisición):
 - **Banda espectral** del sensor: visible, IR térmico, NIR, ..
 - **Tipo de imagen**: 2D, vídeo, 3D, ...
- **Detección** de la cara:
 - Método umbralización color de **piel**.
 - Método de **Viola-Jones**
 - Detección de **landmarks**
- **Reconocimiento** facial:
 - **Extracción de características** y de **matching**.

Detección

Localización de la cara:

- Encontrar las posiciones y delimitar las regiones (Su resultado condiciona el rendimiento del sistema.)
- Dificultad de extraer de forma precisa la cara en imágenes: variabilidad facial en pose, escala, expresiones faciales, color de piel, oclusiones, múltiples personas, variaciones de iluminación

Detector por color de piel: Viola-Jones, Uso de modelos faciales

Métodos para clasificación de piel:

- umbralización en espacios, uso de histogramas y clasificación bayesiana
- Uso de diferentes espacios de color
- capturas en de espectros no visibles (IR).

Detección facial Viola- Jones (2001) un clasificador basado en adaBoost (AdaBoost permite la construcción de clasificadores mediante la combinación de un conjunto de clasificadores lineales “débiles”.) Muchas imágenes de entrenamiento (el cual es lento) pero test muy rápido.

Los landmarks faciales, puntos fiduciales o puntos de referencia son puntos específicos en el rostro que se utilizan para identificar y analizar características faciales.

Reconocimiento

Clasificación de los métodos de reconocimiento facial (Zhao, 2003):

- Holísticos: Usan la región de la cara completa como patrón de entrada al sistema.
- Basados en extracción de características: Primero se extraen ciertos rasgos faciales y después se establecen relaciones geométricas. Con esta información se construyen los vectores de características.
- Híbridos: Combinan la información de los dos métodos anteriores

Técnicas de reconocimiento (Jain et al., 2011):

- Métodos basados en apariencia: Generan una representación compacta de la cara proyectando la imagen a un sub-espacio de dimensión menor (PCA, LDA, ...)
- Métodos basados en modelos: Usan modelos faciales 2D y 3D para facilitar la comparación de imágenes cuando varía la pose
- Métodos basados en texturas: buscan características locales de textura facial que resulten robustas a cambios de pose y de iluminación

Idea PCA, características:

- Color
- Relleno
- Color borde
- Color relleno
- Número lados
- Radio
- Centro

Método Eigenfaces:

1. Recopilar imágenes de entrenamiento que debe: diferentes poses, expresiones y condiciones de iluminación.
2. Preprocesar y normalizar las imágenes para que tengan la misma orientación, tamaño y posición
3. Aplicar análisis de componentes principales (PCA) a las imágenes de entrenamiento para reducir la dimensionalidad y extraer las características más importantes. .
Encontrar los eigenfaces, que son los vectores propios de la matriz de covarianza de las imágenes.
4. Se seleccionan los eigenfaces más importantes
5. Cada imagen de entrenamiento se proyecta en el espacio de eigenfaces
6. Entrenar un clasificador con las proyecciones de las imágenes de entrenamiento en el espacio de eigenfaces
7. Una vez entrenado el clasificador, se puede probar en nuevas imágenes

Filtro de Gabor:

- Filtro lineal para detección de bordes en el dominio de la frecuencia

- Un filtro de Gabor 2D Gabor es una función de tipo kernel gaussiano modulada por una función sinusoidal.

Método Filtro de Gabor para el reconocimiento facial:

1. Adquirir un conjunto de imágenes de entrenamiento.
2. Preprocesar y normalizar su tamaño y orientación
3. Aplicar banco de filtros de Gabor a las imágenes en diferentes escalas y orientaciones:
 1. Construir un banco de filtros de Gabor con diferentes parámetros
 2. Aplicar cada filtro del banco
 3. Calcular el valor absoluto de las respuestas de filtro para obtener una medida de la magnitud de la respuesta y aplicar una operación de pooling, (suma o el promedio), para combinar las respuestas de cada filtro.
4. Seleccionar un subconjunto de características relevantes mediante técnicas de selección de características
5. Entrenar un clasificador utilizando las características seleccionadas y un conjunto de imágenes de entrenamiento etiquetadas.

Método Elastic Bunch Graph Matching (EBGM): Basado en un modelo facial que usa un grafo de información local (textura de regiones de puntos fiduciales (landmark))

Descriptores Local Binary Patterns (LBP): Cada píxel se compara con los vecinos adyacentes y asignar un valor de 1 si el vecino es mayor o igual que el píxel central, o un valor de 0 si es menor. Los valores de 1 y 0 para cada píxel se concatenan en una cadena binaria. Este valor se denomina patrón local binario LBP

Histogram of Local Binary Patterns (HLBP): Se calcula un histograma de los patrones LBP globales en la imagen pero HLBP no representa configuración local

Uniform LBP restringe los patrones LBP a aquellos que tienen al máximo 2 transiciones de 0 a 1 o de 1 a 0

Extended LBP utiliza una ventana de vecinos más grande, como una matriz de 5x5 o 7x7, para calcular el LBP

HLBP con información local se divide la imagen en regiones y se concatena el HLBP de cada región.

Método LBP multi-escala (MLBP):

1. Adquirir un conjunto de imágenes de entrenamiento.
2. Preprocesar y normalizar su tamaño y orientación
3. Calcular histogramas HLBP de las imágenes de entrenamiento:
 1. Usando diferentes radios
 2. La imagen se divide en regiones de las que se calculan histogramas.
 3. Los histogramas determinan vectores de características para el reconocimiento
4. Seleccionar un subconjunto de características relevantes mediante técnicas de selección de características

5. Entrenar un clasificador utilizando las características seleccionadas y un conjunto de imágenes de entrenamiento etiquetadas
6. Una vez entrenado el clasificador, se puede probar en nuevas imágenes para determinar a qué clase pertenecen

Convolutional Neural Networks: utilizan capas de convolución para extraer características de las imágenes de entrada y se combinan con otras capas (activación, agrupación y normalización) para formar una arquitectura de red. Se entrenan para extraer automáticamente características relevantes de las imágenes

Arquitecturas Siamesas CNN: Utiliza dos redes idénticas que procesan dos entradas diferentes para generar representaciones vectoriales de cada entrada que posteriormente compara en una capa de comparación. Mide la similitud entre los vectores. El objetivo es ajustar los pesos de manera que los vectores sean lo más similares posible para entradas similares. Una vez entrenada, se puede utilizar solo una rama.

Reconocimiento 3d

Reconocimiento facial 3D: hace uso de información tridimensional de la forma y la textura facial obtenida mediante p.ej. un escáner 3D. Tipos de info:

- Nube de puntos
- Modelos Maya
- Mapas de profundidad

Tecnologías de adquisición:

- Láser Se utiliza un rayo láser para medir la distancia entre el objeto y el escáner
- Luz estructurada proyectan patrones de luz sobre un objeto y miden la deformación de los patrones
- Fotogrametría - Imágenes estéreo imágenes 2D tomadas desde diferentes perspectivas para calcular por triangulación la posición
- TOF (Time of Flight) teniendo en cuenta el tiempo que tarda una señal de luz desde el sensor al objeto y volver al sensor
- Tomografía computarizada : la tomografía computarizada utiliza rayos X para capturar imágenes 3D

Ventajas:

- Mejora de la precisión
- Eficiencia
- Robustez

Desventajas:

- Dispositivos de captura

- Más costoso en términos de tiempo de cómputo y memoria
- Dispositivos más costosos
- Bases de datos

Multimodal 2D + 3D:

- Mejora el resultado
- Se crea un vector de características que contiene datos de la imagen 2D y también datos de la 3D
- No se puede asumir la independencia entre los datos

Reconocimiento facial en video:

- Los datos de entrada son secuencias de vídeo
- Es necesario seguir las caras.
- Necesita una buena detección inicial
- Técnicas:
 - Set-based: Consideran los vídeos como colecciones no ordenadas
 - Sequence-based: Usan la información temporal para mejorar el reconocimiento

Ventajas:

- Robustez ante cambios de iluminación, pose, expresiones u oclusiones faciales
- Dinámica temporal
- Se puede inferir información 3D

Desventajas:

- Estos métodos son más costosos computacionalmente.
- Se incrementa la variabilidad intra-clase

Reconocimiento facial multiespectral:

- La piel y el rostro humano emiten una radiación térmica
- Infrarrojo cercano vs infrarrojo térmico
- Las variaciones de temperatura en la superficie de la cara producen un mapa del calor llamado termograma
- Debido a la presencia de características distinguibles y permanentes en la piel facial, los termogramas contienen información importante que puede ser usada para el reconocimiento facial.
- No es posible extraer buenos descriptores de textura.
- Técnicas similares a las usadas en el espectro visible
- Multimodalidad

Ventajas: Invariante a cambios de iluminación.

Desventajas: Precio elevado

Huellas dactilares

Morfología

Es un patrón morfológico caracterizado por la presencia de un conjunto de líneas genéricas llamadas crestas (ridges) que corresponden a zonas donde la piel se eleva sobre unas zonas bajas o valles (valleys).

Dactilograma: figura formada por el relieve de las crestas papilares. Zonas:

- Base: parte inferior
- Margen: contorno lateral y superior
- Núcleo: parte central, entre base y margen

Minucias (crestas): puntos característicos en los que las crestas papilares se bifurcan, interrumpen y/o unen. Puntos característicos:

- Puntos de terminación de crestas (endings)
- Puntos de bifurcación de crestas (bifurcations)

Otras características: puntos core, delta, poros, pliegues, cortes...

Clasificación de huellas:

- Henry: "las crestas papilares se agrupan para formar sistemas definidos por la uniformidad de su orientación y figura".
- Número de deltas presentes: adeltas (20% de huellas), monodeltas (50%) y bideltas (30%)

Tipos de huellas:

- Ejemplares: aquéllas capturadas de forma deliberada y en condiciones controladas (Huella completa y limpia)
- Latentes: impresiones accidentales invisibles (huellas parciales, distorsionadas y con ruido)
- Patentes: impresiones accidentales visibles
- Plásticas: realizadas sobre un material que captura el relieve

Biometría dactilar:

- Dactiloscopia: estudio analítico de las huellas dactilares
- La huella dactilar se manifiesta a partir del 6º mes del embrión como consecuencia de un proceso aleatorio, no genético
- Modalidad biométrica por excelencia, más tradicional, presente en pasaportes y muchos documentos de identidad.
- Constituyen en torno a la mitad del mercado biométrico

Ventajas:

- Son invariantes al paso del tiempo
- Fácil adquisición y amplia aceptación.
- Alta unicidad y la permanencia del rasgo
- Buenos algoritmos con muy elevadas tasas de éxito (99.4% de TAR al 0.01% de FAR en BD del NIST)
- Existencia de lectores de huellas precisos, compactos y baratos.
- Existencia de importantes BD de huellas.

Desventajas:

- dificultades para la captura:
 - Colectivos étnicos y personas de edad avanzada
 - Colectivos profesionales: algunas profesiones manuales y personas que trabajan con productos químicos.

- Problemas en la captura (online vs offline)

Otras modalidades relacionadas: Huellas de la planta del pie (footprints) y huellas palmares (palma de la mano)

Captura

Es la etapa más crítica y determina la calidad de la imagen de la huella con una influencia importante en el rendimiento del sistema.

Captura off-line (entintado del dedo) frente a la captura on-line (live-scan inkless).

Adquisición multi-dedo vs mono-dedo:

- Multi-dedo: Se capturan varios dedos de la mano.
- Mono-dedo: Solo se captura un dedo simultáneamente

Tipos de lectores (scanners) live-scan de huellas: ópticos, capacitivos, ultrasonidos, térmicos, ...

Dispositivos ópticos:

- FOTOSENSORES:
 - Ventajas: Elevada resolución
 - Desventajas: sensibilidad a la suciedad y aberración en la zona marginal

Dispositivos capacitivos: Sensor electromagnético (traduce a niveles de gris la diferencia de capacidades):

- Ventajas: bajo consumo y reducido tamaño
- Inconvenientes: sensibilidad a la humedad de la huella y contaminación electromagnética

Tecnología de ultrasonidos: ondas ultrasónicas:

- Ventajas: mas calidad de imagen, resistente ataques y mas resistente a suciedad, sudor
- Inconvenientes: mayor precio y menor resolución de la imagen.

Touchless: factibilidad de conseguir huellas a una cierta distancia:

- Mejorar tasas de reconocimiento/verificación automática con imágenes de baja calidad.
- Técnicas de fingerprint mosaicing (combinar pequeñas porciones de huellas)
- Automatización en el proceso de análisis de huellas latentes.

Etapas de preprocesado de la imagen: Al ser la captura imperfecta no es directo determinar los píxeles de la imagen que pertenecen a crestas. Algunos pasos usuales son:

- Mejora (filtrado y realzado) de la imagen para eliminar redundancias: filtros direccionales para disminuir ruido y aumentar el contraste
- Estimación del campo de orientación local de las crestas: se divide la imagen en bloques (15x15), se calcula el gradiente en x e y de cada píxel, y a partir del gradiente se calcula su orientación de cada bloque
- Binarización: se convierte la imagen en bitonal (algoritmo de Otsu)
- Adelgazamiento (thinning): convierte el grosor de las líneas de las crestas a un solo píxel (algoritmo de esqueletizado)
- Depuración: elimina las ramas parásitas en las crestas

Depuración = Post-proceso para eliminar falsas minucias.

Reconocimiento/Verificación: técnicas para comparar dos huellas:

- Huella completa : el uso de técnicas generales de reconocimiento de patrones.
- Puntos minucias : se comparan las posiciones y orientaciones de las minucias

Etapas de extracción de características:

- Obtener la localización, la orientación y el tipo de minucias.
- Solo se almacenan datos correspondientes a la disposición de minucias
- vectores de tamaño variable.
- Localización de terminaciones y bifurcaciones de crestas mediante análisis de los 8 vecinos de cada píxel

Etapas de reconocimiento (matching):

- Una vez alineados, busca una minucia de referencia y se ordenan las minucias de los dos modelos
- Se efectúa una comparación basada en la distancia euclídea entre los vectores test y modelo

Iris

Biometría del Iris: es una membrana pigmentada suspendida en el interior del ojo, entre la córnea y el cristalino. Regula el tamaño de la pupila para controlar la cantidad de luz que entra al ojo.

Patrón: Características aleatorias de color y textura en el músculo del iris.

Ventajas:

- Elevada permanencia y universalidad (estable desde el octavo mes de gestación)
- No puede ser alterado fácilmente como otras biometrías (protección de la cornea). Bajas posibilidades de ser atacada (circunvención).
- grado de discriminación de esta modalidad elevadísimo (gran aleatoriedad en el patrón)
- Alta tasa de reconocimiento

Desventaja:

- Mayor grado de intrusión en la captura

Captura: normalmente, con iluminación en espectro NIR. Tres tipos de dispositivos:

- Tradicional: Imagen de alta calidad, pero sistema poco flexible y amigable
- Media distancia: Sistema LG iCAM4000 (más flexible y amigable)
- Larga distancia: Iris-On-theMove™ (IOM) de Sarnoff Corporation (alta flexibilidad, usa 4 cámaras IR de alta resolución)

Segmentación del iris:

- Región incluida en dos círculos concéntricos: uno indica la separación entre el iris y la pupila, y la otra dentro del iris. Se usa la textura en la región entre esos dos círculos.
- Método de Wildes:
 - Aplicar un detector de bordes
 - Encontrar exactamente los dos círculos más importantes con el mapa mediante un mecanismo de votación (transformada de Hough)

Extraer características de la textura del iris y almacenarlas de forma compacta. Uso de bancos de filtros de Gabor (J. Daugman): • Uso de la distancia de Hamming normalizada para comparar representaciones binarias de dos iris.

Líneas de trabajo:

- Buscar que sea menos intrusiva, más aceptada y popular
- Con el desarrollo se realizarán de forma más rápida en dispositivos de tamaño más reducido
- Reconocimiento multi-espectral
- Localización “menos controlada”

- Procesamiento de iris “con ruido”
- Multi-modalidad biometrías oculares.

Retina

Biometría de la retina:

- Vasos sanguíneos de la retina
- El escáner emite una luz infrarroja que se refleja en la retina y se recoge de nuevo por el escáner

Ventajas:

- patrón que presenta unas características únicas para cada persona
- hasta más de 400 puntos únicos
- se puede codificar usando solo 96 bytes.

Desventajas:

- Captura muy intrusiva
- El matching entre dos retinas implica definir una sistema de puntuaciones entre codificaciones y definir una medida de similitud entre los patrones.
- Pocos trabajos publicados.

Geometría mano

Geometría de la mano:

- se basa en la forma del contorno de la mano
- procede del sistema de Bertillon (s. XIX).
- Se la considera una biometría de “perfil medio” con un grado de aceptación alto y un coste computacional bajo.
- Las tasas de error son mayores que las obtenidas por las biometrías de “alto rendimiento”
- Usada más para verificación.
- Algunos investigadores han incluido esta tecnología como parte de un sistema biométrico multimodal

Sistema de verificación/identificación:

- Eliminación del fondo: $(IR+IG)-IB$.
- extraer el contorno de la mano (Filtro de Sobel)
- Umbralización para obtener una imagen binaria
- Se extraen un conjunto de medidas del contorno

- Posible aplicación de PCA para reducir la dimensionalidad.
- Clasificación: distancia euclídea, distancia Hamming, modelos basados en mezcla de gaussianas (GMM),...
- Mejoras: sistemas “pegs-free” (sin clavijas) y sistemas sin contacto.
- Nuevos modelos y clasificadores: Active Shape Models, Bsplines, ...

Huella palmar

La huella palmar (imagen de la palma de la mano) es rica en información que resulta ser única y estable.

Características extraíbles de la huella palmar: líneas principales, pliegues, puntos de minucias o textura.

Ventajas:

- Comparadas con las huellas dactilares, la palma presenta mayor superficie y permite extraer más características.
- Las líneas de la mano son estables durante toda la vida (menor probabilidad de dañarse)
- Menos suciedad en la palma que en los dedos
- se puede trabajar a más baja resolución
- Menos problemas con la iluminación (se trabaja con iluminación controlada)
- Multimodalidad

Desventajas:

- Dificultades ante suciedad de la mano y sensores.
- Ruido de captura, deformaciones,..
- Los resultados se presentan sobre BBDD reducidas
- Dificultad “real” de esta modalidad: uso de anillos de diferentes tipos y tamaños.

Vasos sanguíneos

Biometría de vasos sanguíneos:

- Imágenes térmicas de los patrones de la venas del dorso de la mano.
- Se extrae la ROI y después un conjunto de puntos característicos (Tasas de error en torno al 2% (FAR y FRR) para verificación)
- Técnicas similares se han aplicado para la red venosa de los dedos.
- Usa un dispositivo que captura el patrón de la imagen en el infrarrojo cercano.
- Imágenes “no-claras”
- Trabajo posterior de Parken2011 (EER del 1,53%.)

Oreja

Biometría de la oreja:

- muchas características distintivas interpersonales: su localización, sus medidas y orientaciones.

Ventajas:

- La oreja es un rasgo bastante estable que no cambia mucho su morfología durante la vida humana
- no son del todo simétricas y su crecimiento es proporcional con el del cuerpo desde los 4 años.
- Representa aproximadamente 1/20 o 1/25 de la imagen facial de perfil
- Resultan ser patrones más “seguros” que la imagen facial porque es muy difícil asociar la imagen de la oreja con una persona

Desventajas:

- Su oclusión por el pelo y por pendientes dificultan la extracción.
- Problemas iluminación

Dentadura

Biometría dental:

- Utiliza información sobre las estructuras dentales para identificar individuos.
- Patrones correspondientes a imágenes de radiografías dentales
- Se aplica principalmente en la identificación de personas en accidentes y desastres naturales
- Se miden características de las piezas dentales extraídas de radiografías (características morfológicas de las piezas y también de su aspecto (niveles intensidad)).

Movimiento de labios

Biometría Movimiento de labios al decir frases, palabras:

- Uso de secuencias de vídeo
- Extracción de la región de los labios
- Búsqueda de puntos característicos y extracción de relaciones entre puntos
- Extracción de características basadas en movimiento
- Multimodalidad con voz
- Tasas de reconocimiento de un 89%

Soft Biometrics

Uso de “soft biometrics”: (de forma aislada o en combinación con otras modalidades)

Otras modalidades fisiológicas

- **Reconocimiento por la sonrisa (smile recognition):** Guan et al, 2004
- **Análisis de los movimientos oculares al leer**
- **Textura de nudillos en los dedos:**
- **Biometría de las uñas (nail ID)**
- **Electrocardiograma(ECG), electroencefalograma(EEG)**

Tema 5. Reconocimiento de Patrones

Introducción

Reconocimiento de patrones: disciplina que diseña algoritmos que buscan regularidades en los datos con el fin de agruparlos en clases o categorías.

Casos de uso:

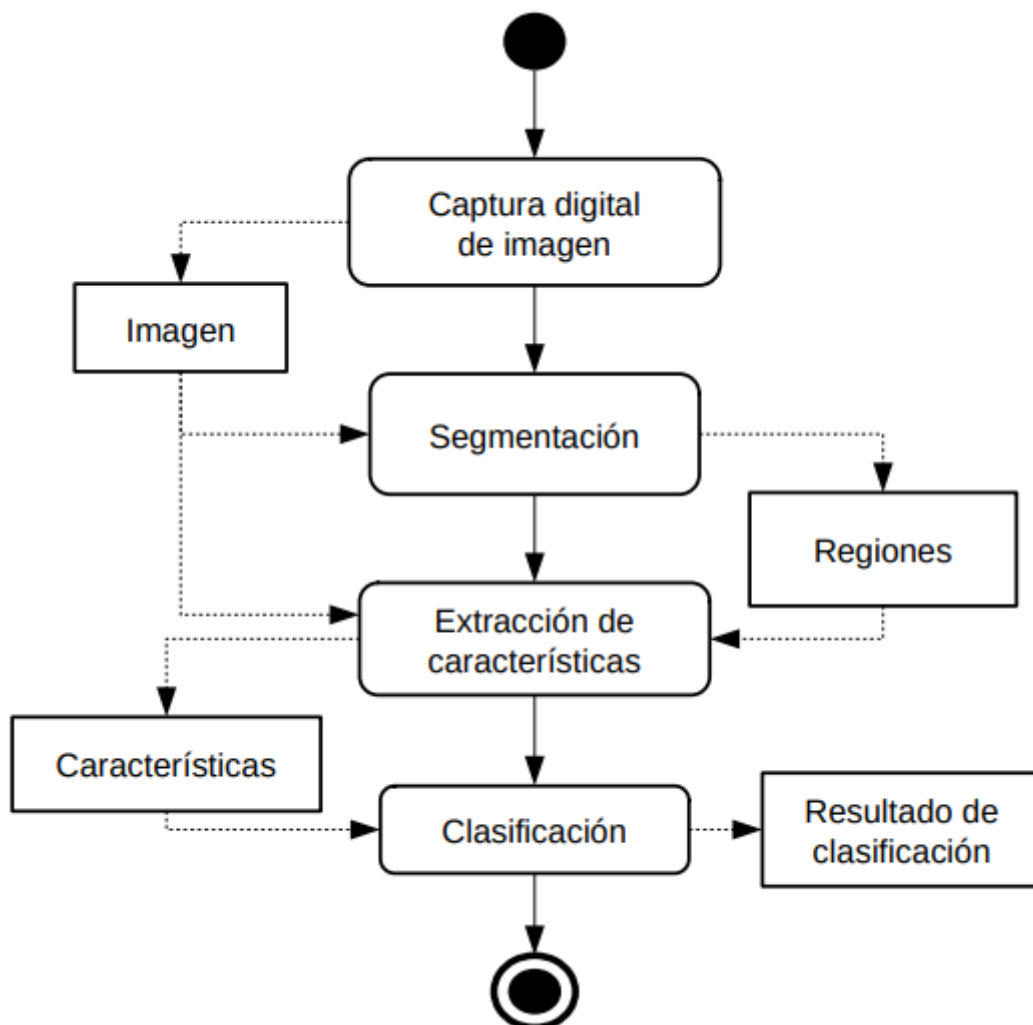
- Reconocimiento de objetos, lugares, expresiones etc.
- Inspección de objetos manufacturados
- Reconocimiento de voz
- Diagnóstico clínico
- Predicción de catástrofes
- Clasificación de textos (detección de Spam...)

Aplicaciones:

- Clasificación de objetos: asocia la imagen a una clase determinada
- Detección y Reconocimiento: localiza dentro de la imagen el objeto, y lo clasifica en una categoría
- Segmentación Semántica: Busca regiones de interés.
- Segmentación de Instancias: Trata a distintos objetos como categorías independientes.
- Segmentación Panóptica: Segmenta la imagen en partes o regiones semánticamente significativas, al mismo tiempo que detecta y distingue instancias individuales de objetos dentro de estas regiones.
- Imágenes Médicas:
 - Ayuda al diagnóstico
 - Apoyo en operaciones quirúrgicas
- Face Recognition: Identificación, Análisis Expresiones, Filtros, Ver Tema Biometría
- Estimación Pose: Seguimiento, Realidad aumentada, Robótica
- Estimación Pose Densa
- Transferencia de Estilo

- Deep Fake
- Búsqueda de Información (Information Retrieval)
- Generación de Imágenes a partir de texto
- Descripción de Imágenes
- Reconocimiento de Texto
- Verificación firmas
- Identificación escritor
- Transferencia de estilo en textos
- Indexación de palabras (word spotting)

Fases en habituales de un sistema de VA:



Intuición de problema de clasificación: Para distinguir las imágenes de un 0 y un 1 comenzamos por abstraer algunas características importantes de la imagen en forma de vector de características numérico:

- Enfoque holístico: Las características son las propias imágenes de entrada
- Características de alto nivel: Son generadas por manipulaciones, mediciones o transformaciones de las imágenes.

Intuición muestra de aprendizaje: Luego obtenemos esas características para un conjunto de imágenes de aprendizaje.

Intuición de clasificación de patrones: Finalmente, creamos una función $f(x)$ que permita clasificar dichos vectores de características como pertenecientes a una clase u otra.

Conceptos:

- El Universo de trabajo (UT) es el conjunto de objetos disponibles para la construcción de un clasificador.
- Una clase es una agrupación de objetos del universo de trabajo que comparte ciertas características (ej, En un problema de reconocimiento de números manuscritos las clases serían $\{0, 1\}$, En un problema de identificación de personas las clases podrían ser $\{\text{Andrés, Juan, Pedro...}\}$)
- Un objeto o patrón es una instancia o ejemplar de una clase (ej, En el caso de reconocimiento de números manuscritos un patrón sería cualquier imagen de un cero o de un uno., En el caso del reconocimiento de personas, un patrón sería cualquier fotografía de una persona.)
- Un vector de características discriminantes es un descriptor numérico en \mathbb{R}^n que representa un objeto y facilita clasificarlo (ej, En el caso de reconocimiento de números que se ha presentado antes sería un par de números reales, En el caso de imágenes de la cara podrían ser medidas entre puntos importantes (distancias entre ojos, de nariz a barbilla...))
- La Función Discriminante es una función que combina las componentes del vector de características y nos permite decidir a qué clase pertenece un objeto. Tipos:
 - Basadas en regionalización.- El espacio de características se particiona, y los patrones se clasifican según la partición en la que se encuentran.
 - Basadas en distancia.- Se define una función de distancia que a cada patrón asigna una medida de pertenencia a cada una de las clases. Que a su vez se dividen en:
 - Paramétricos
 - No paramétricos

División de la muestra: Se tendrán en cuenta tres conjuntos de datos DISJUNTOS:

- Conjunto de entrenamiento (TRAIN): usado para calcular la función discriminante.
- Conjunto de validación (VALID): usado para comprobar si la función discrimina bien con datos nuevos
- Conjunto de evaluación (TEST): usado para ver si la función es generalizable (si funciona bien con datos nuevos)

Tipos de clasificadores:

- A priori: Se construyen en un solo paso.
- A posteriori: Se construyen mediante un proceso de optimización que se denomina aprendizaje.
- Deterministas: El clasificador asigna a cada patrón la pertenencia a una única clase.

- No deterministas: El clasificador asigna a cada patrón un grado de pertenencia a cada clase.
- Supervisados: cuentan con datos de entrada y datos etiquetados de salida (ground truth).
- No supervisados: solo utilizan datos de entrada (son siempre no paramétricos)
- No paramétricos: aquellos en los que no se conoce la distribución (o modelo) de los datos de cada clase
- Paramétricos: aquellos que se conoce la distribución (o modelos), ej regresión lineal

Estimación del rendimiento

Objetivo: Preferiremos aquél que mejor generalice: el que mejor clasifique muestras que no se emplearon para entrenarlo.

Cómo medir el error: La Tasa de Error mide la capacidad generalizadora de un clasificador: $\frac{\text{ErroresEnLaClasificacion}}{\text{muestrasTotales}}$ Idea:

1. Se va a considerar una clase de referencia.
2. Se va a considerar un contraste de hipótesis donde la hipótesis nula es H_0 : "La imagen contiene un objeto de esa categoría"
3. Si la imagen contiene un objeto de esa categoría se acepta la hipótesis nula
4. Si la imagen no contiene un objeto de esa categoría se rechaza la hipótesis nula

A la hora de clasificar pueden suceder las siguientes 4 situaciones:

Realidad/ Resultado	1 (aceptar)	0 (rechazar)
1 (aceptar)	True Positive (TP), True Accepted (TA)	False Negative (FN), False Rejected (FR) Error tipo II
0 (rechazar)	False Positive (FP) False Accepted, FA) Error Tipo I	True Negative (TN) True Rejected (TR)

• Si voy a cometer algún error, ¿qué es peor? ¿Tener un FP o un FN?

• En estadística, el **error tipo 1 se considera más grave** y por ello intenta minimizarlo.

Ratios directos:

- FNR, FRR, error Tipo I, error al rechazar: $\text{FN} / (\text{FN} + \text{TP}) = \text{FN} / P$
- FPR, FAR, error Tipo II, error al aceptar: $\text{FP} / N = \text{FP} / (\text{FP} + \text{TN}) = \text{FP} / N$
- TNR, TRR, acierto al rechazar, especificidad: $\text{TNR} = \text{TN} / (\text{FP} + \text{TN}) = \text{TN} / N$

- TPR, TAR, acierto al aceptar, recall, sensibilidad, exhaustividad: $TPR = TP / (FN+TP) = TP / P$

Otros ratios:

- Accuracy, Exactitud: $(TP+TN) / (TP+FP+TN+FN) = (TP+TN) / T$
- Precision, Precisión, Positive Predictive Value: $TP / (TP+FP)$
- Índice de Jaccard, IoU: $TP / (FN+TP+FP)$
- F1 Score: media armónica precision y recall: $2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$
 $2TP / (FN+2TP+FP)$

Ejemplo:

¿Cuál es la precisión, exhaustividad y F1-score de los siguientes problemas?

- Si en una escena hay 9 perros y varios gatos, y nuestro sistema detecta 7 perros, 4 que corresponden a verdaderos perros y 3 que corresponden a gatos:
 - La precisión es de 4/7 y la exhaustividad de 4/9.
F1-score es $2 \times (4/7) \times (4/9) / ((4/7) + (4/9)) = 1/2$
- Si al reconocer 100 cheques, se reconocen bien 80 y se rechazan 20. Y de las 80 reconocidas, 60 se aciertan y 20 se fallan.
 - La precisión es de 60/80 y la exhaustividad 60/100.
F1-score es $2 \times (60/80) \times (60/100) / (60/80 + 60/100) = 2/3$

El error está en función de un parámetro:

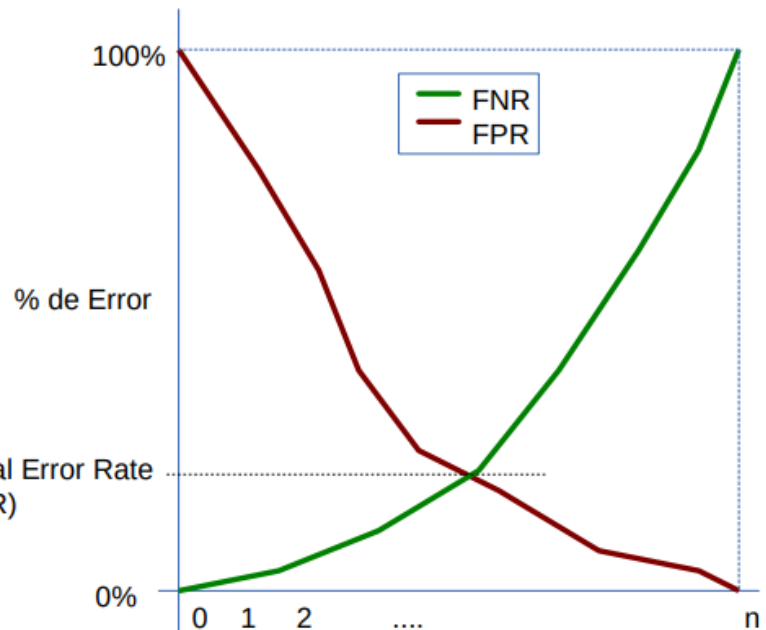
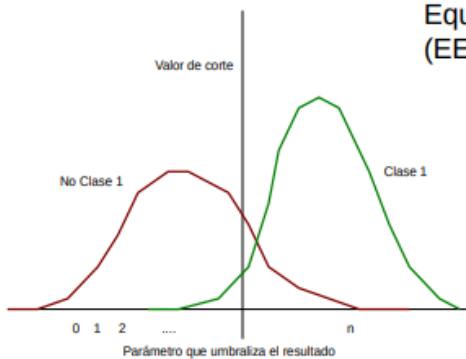
- Muchos clasificadores, en lugar de asignar directamente una instancia a una clase, dan la confianza ("probabilidad") de que esa instancia pertenezca a cada una de la clase.
- Hay que decidir, a partir de qué confianza se asgina a una clase u a otra.

Equal Error Rate:

Muestra FNR y FPR.

$$FNR = \int_{-\infty}^U \text{Clase 1}$$

$$FPR = \int_U^{\infty} \text{NoClase 1}$$



Parámetro que umbraliza el resultado

58/161

Matriz de Confusión: Tabla de frecuencias que recoge los resultados del modelo. De esta tabla se pueden obtener los TP, TN, FP, FN y todas las medidas que derivan de ellos.

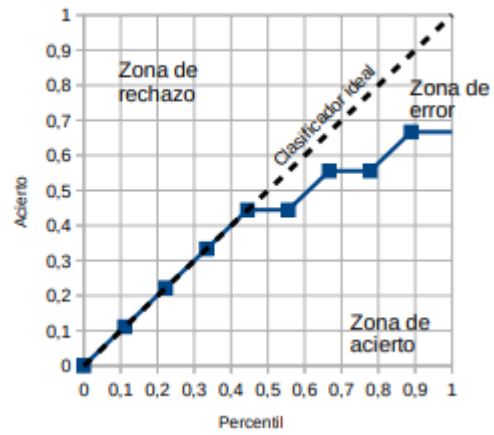
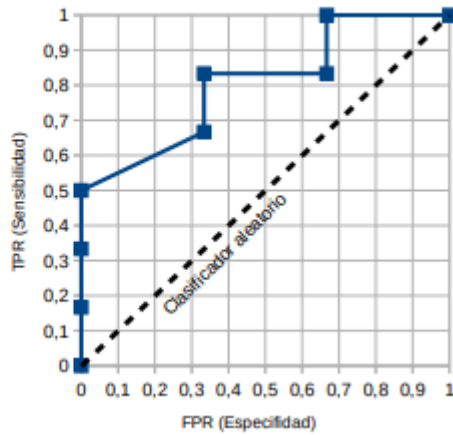
Curva ROC: El espacio Receiver Operator Characteristic (ROC) informa sobre los aciertos capturados (sensibilidad o recall) respecto a los falsos positivos proporcionados por el clasificador según cambia el umbral de rechazo (aplicable a problemas verificación):

- Permite comparar en la misma gráfica a varios clasificadores. Cuanto más cercanos a la esquina superior izquierda mejor.
- El Área Bajo la Curva (AUC) es una medida de lo “bueno” que es el clasificador.

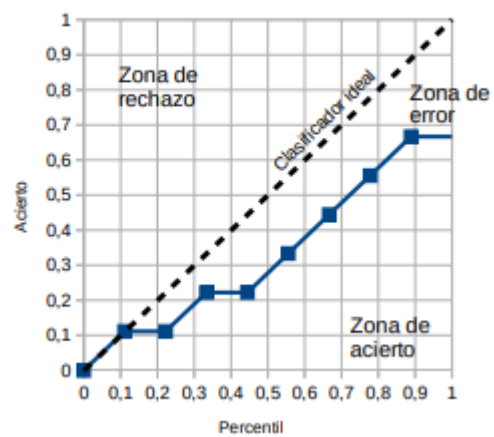
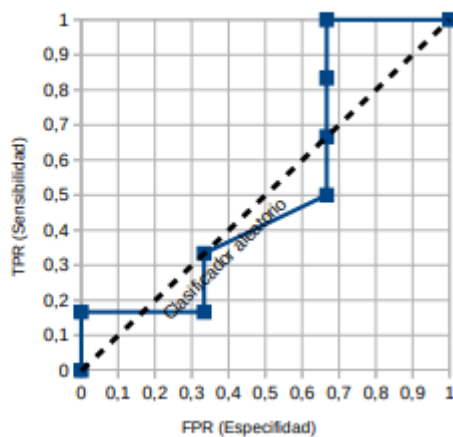
Gráfico de Ganancia Acumulada (CG): La curva Cumulative Gain (CG) informa sobre la tasa global de acierto (accuracy) facilitando ver las posibles relaciones de acierto, rechazo y error que ofrece un clasificador. Es aplicable a clasificadores binarios (aplicable a clasificadores binarios)

Comparativa entre las curvas presentadas:

A	A
A	A
A	A
A	A
A	B
A	A
A	B
A	A
A	A
A	B



A	A
A	B
A	A
A	B
A	A
A	A
A	A
A	A
A	A
A	B



Clasificadores Supervisados

Clasificador de distancia euclídea

Hipotesis del clasificador euclídeo: Este clasificador es adecuado cuando se supone que la dispersión (σ) de las clases es pequeña en relación a la distancia (d) entre ellas. Idea: Calcular la distancia de un nuevo punto a cada clase, y elegir la que esté más “cerca”.

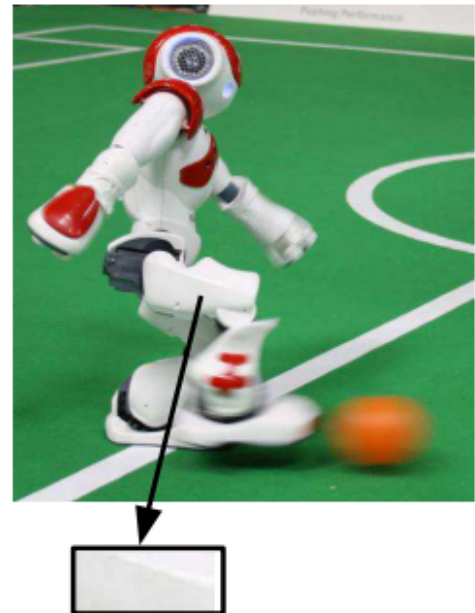
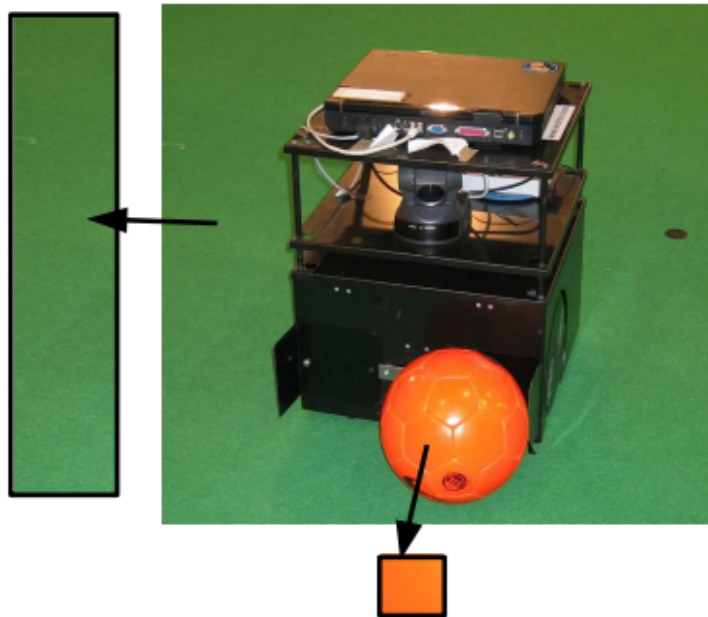
Cálculo de centroides: El clasificador euclídeo necesita que para cada clase se defina un representante, o centroide, que se calcula como la media de los vectores de características de los patrones que existen para esa clase en el conjunto de aprendizaje.

Clasificación:

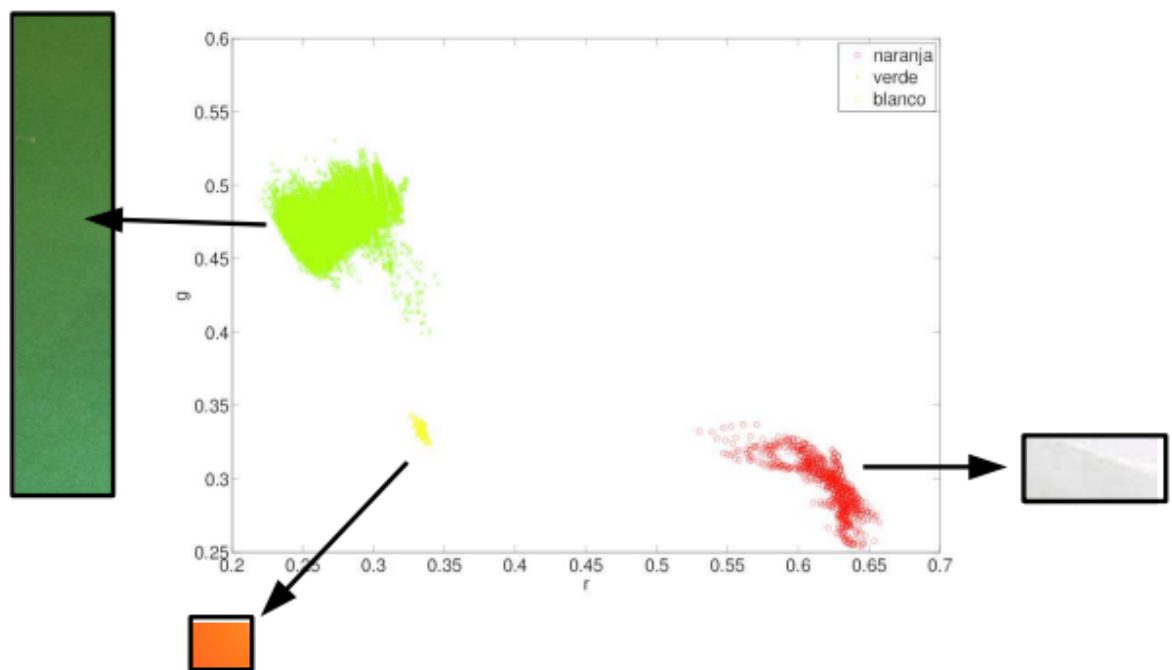
- Un patrón se clasifica en la clase cuyo centroide está más cerca utilizando la distancia euclídea.
- El clasificador euclídeo genera una separación lineal entre las clases que separa.
- Se puede usar para un número indefinido de clases, y siempre genera separaciones lineales
- Se puede usar para un número indefinido de clases.

Ejemplo de clasificación por color de píxel:

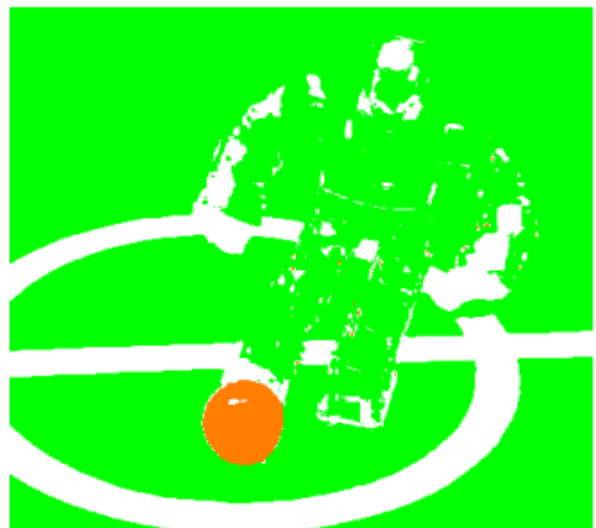
Necesitamos extraer datos de entrenamiento
(regiones de la imagen):



Espacio de características:



Resultados de la clasificación sobre imagen parecida



K-vecinos más cercanos

Un punto en el espacio es asignado a la clase α si esta es la clase más frecuente entre los k ejemplos de entrenamiento más cercanos (Generalmente se usa la distancia euclidiana).
Idea: Elegir los k puntos más “cercaños”, y elegir la clase más frecuente de esos puntos.
En 2 dimensiones da lugar a una segmentación del espacio en células de Voronoi.

Support Vector Machine (SVM)

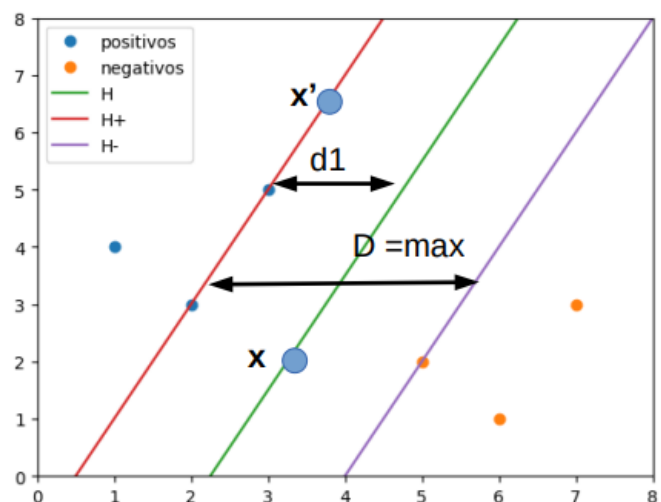
Máquina Vector Soporte Lineal: El problema que se plantea es el siguiente: identificar un sistema automático de clasificación. Una forma de abordar este problema consiste en representar un hiperplano que separe las dos. Idea: Plano que maximice la separación entre las clases

Modelo de optimización: Como quiero maximizar tendré que resolver un problema de optimización. Hay que definir:

- Variables: Variables: w, b Dado un conjunto de puntos clasificados, se quiere encontrar el hiperplano $w^t x = b$ que separe al máximo las clases "+1", "-1".
- Restricciones: Se distinguen las dos clases de puntos:
 - $(w^t x_i - b) > 0$
 - $(w^t x_i - b) < 0$
 - Se multiplican las restricciones por λ , de tal manera que el punto más cercano al hiperplano esté a distancia 1:
- Función objetivo: Entre todos los hiperplanos que separan las dos clases, se supone que son separables, como en el ejemplo 1, hay que elegir el que deja un margen entre ambos máximo.

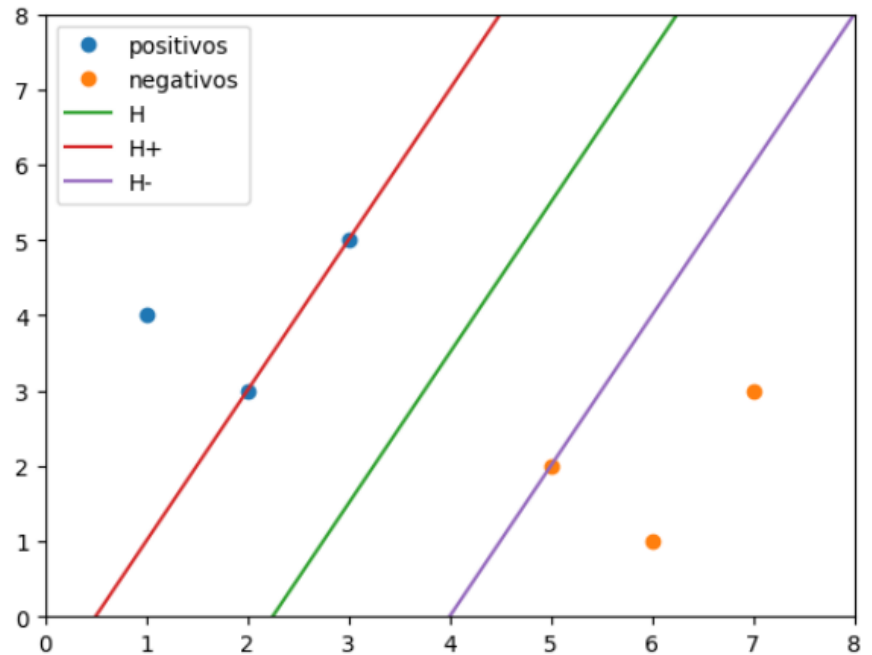
Sea x un punto de H y x' un punto de H^+

Maximizar la distancia entre H y H^+ equivale a minimizar el cuadrado de la distancia entre los dos hiperplanos



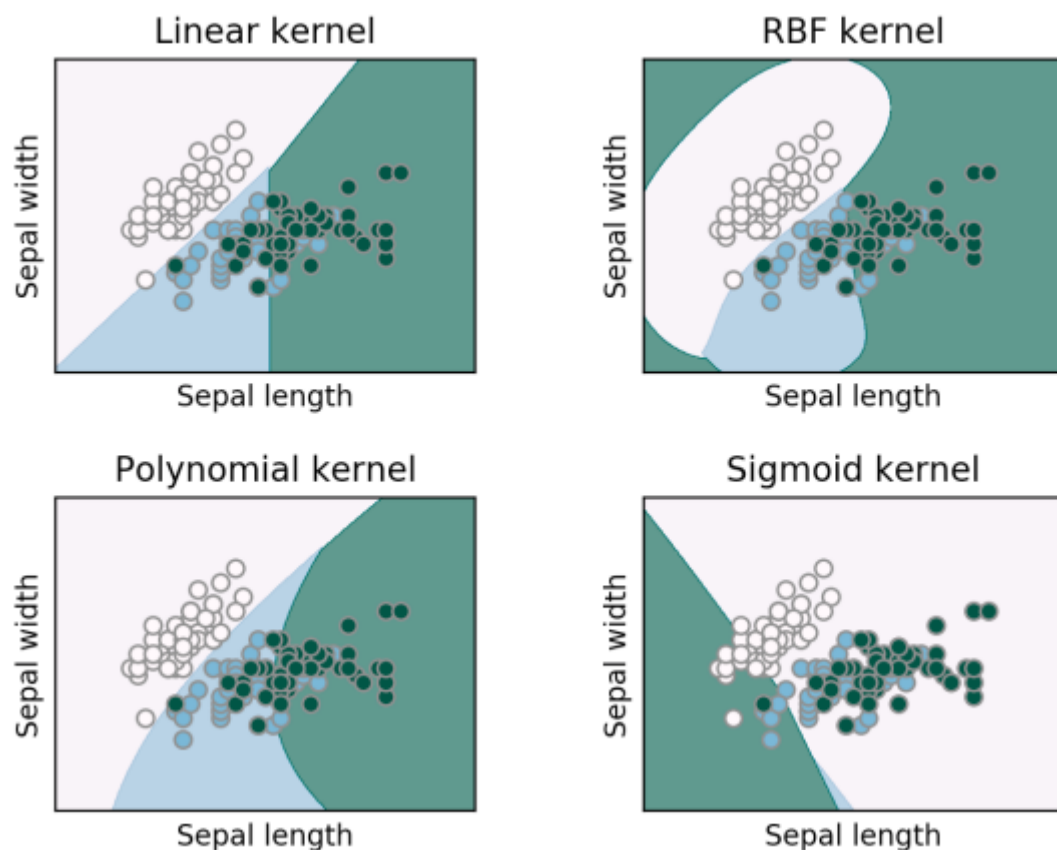
Resolviendo el problema, se tiene el hiperplano :

$$4x_1 - 2x_2 = 9$$



Generalizando se ha encontrado a partir de los datos x una medida de similaridad k (kernel) que permite la clasificación de un nuevo punto z .

Otras Máquina Vector Soporte: Si cambiamos la función kernel k , se tendrán otro tipo de regiones de separación:



Clasificadores No Supervisados

K-medias

Introducción: Dado un conjunto de observaciones $\{x_1, x_2, \dots, x_n\}$ donde cada observación es un vector real de d dimensiones, k -medias (k -means) construye una partición de las observaciones en k conjuntos ($k \leq n$) $S = \{S_1, S_2, \dots, S_k\}$ donde se cumple

$$\arg \min_s \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

que: donde μ_i es la media de puntos en S_i .

Idea: Agrupar los puntos más parecidos según la media.

Algoritmo:

Paso 1.- Se hace $t = 1$. Se toman al azar k vectores de los P existentes y se convierten en centroides de cada una de las k clases respectivamente.

$$\begin{aligned} &Z_1(1) \text{ de } 1 \\ &\dots \\ &Z_k(1) \text{ de } k \end{aligned}$$

Paso 2.- Se distribuyen las P muestras entre las k clases. Se asigna cada vector x a la clase cuyo centroide esté más próximo.

$$x_j \text{ } x - z_j(t) < x - z_i(t) \quad \forall i = 1, 2, \dots, k \text{ } i \neq j$$

Paso 3.- Se calculan los centroides de las clases como la media ponderada de los vectores de cada clase.

Paso 4.- Si alguno de los k centroides $Z_k(t)$ es distinto de los nuevos centroides $Z_k(t+1)$ hacer $t = t+1$ e ir al paso 2, en otro caso finaliza el algoritmo.

Reducción de la dimensionalidad

Dimensión del vector de características: Normalmente se crean vectores de características de elevado tamaño. Estos vectores tienen varias desventajas:

- Problema de la dimensionalidad.
- Mayor coste computacional.
- Dificultad para visualizar los datos.

Con objeto de reducir la dimensión hay dos tipos de técnicas:

- Técnicas de eliminación de características
- Técnicas de transformación de las características: Aplicar una función matemática que transforme los datos reduciendo la dimensión para facilitar la clasificación.

Técnicas lineales más importantes:

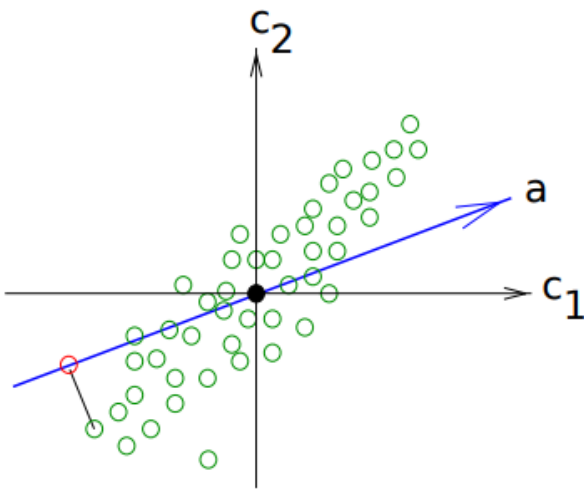
- El Análisis de Componentes Principales (PCA)
- El Análisis Discriminante Lineal (LDA)

Hay otras técnicas no lineales, pero se salen del alcance de este curso:

- PCA suele utilizarse en situaciones en las que no se conoce las clases a las que pertenecen los datos de muestra (no supervisado)
- LDA suele utilizarse cuando se conocen las clases (supervisado).

PCA

Análisis de Componentes Principales (PCA): es una técnica que maximiza la varianza de las características discriminantes. Consiste en encontrar la dirección que maximiza la varianza de los datos o que minimiza la suma de los errores de reconstrucción:

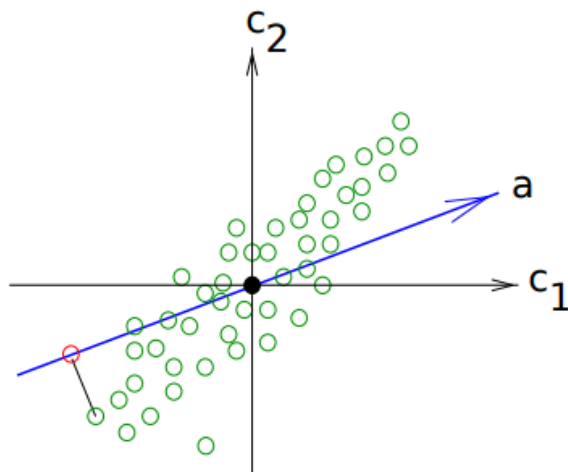


IDEA: Diagonalizar Σ

Los autovectores son las componentes principales. Representan las direcciones de proyección de mayor variabilidad

Los autovalores indican el porcentaje de variabilidad que recoge el correspondiente autovector.

Enfoque descriptivo



Generalizando, se trata de encontrar la matriz A tal que:

$$A_{k \times d} \cdot X_{d \times n} = Z_{k \times n}$$

PCA encuentra la matriz A **maximiza la dispersión de las características en Z :**

$$\frac{A^t \Sigma_X A}{\|A\|^2}$$

114/16

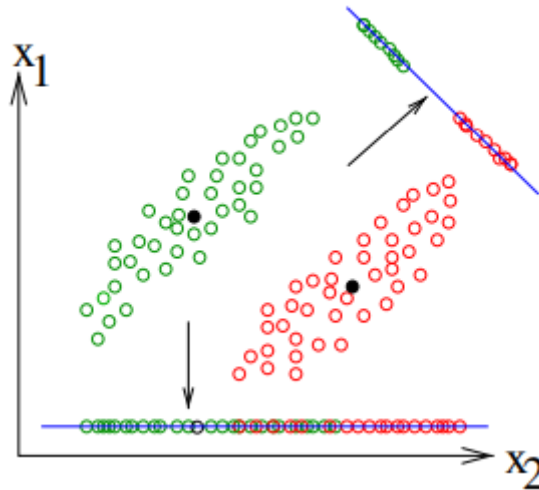
Algoritmo PCA:

- Entrada: Matriz X con m vectores de características.
 - Salida: Matriz de proyección A y la varianza de cada dirección V .
1. Calcular la media de todas las características X y restársela a todas las columnas de X .
 2. Calcular C como la matriz de covarianza de R .
 3. Calcular los autovalores A y autovectores V de C .
 4. Devolver A y V

Selección de las componentes: La reducción de dimensión que se consigue con PCA consiste en quedarse con un subconjunto de los autovectores, y Es usual quedarse con el menor número de autovectores que retengan un porcentaje fijo de la variabilidad.

LDA

Análisis Discriminante Lineal (LDA): técnica que maximiza un índice de separabilidad de las clases en el espacio transformado. Intuitivamente, busca las direcciones que permiten



obtener mayor capacidad discriminante:

Cociente de Fisher: Sean dos clases (α, β) , con medias (μ_α, μ_β) y desviación típica $(\sigma_\alpha, \sigma_\beta)$. El cociente (o ratio) de Fisher para una característica i , se define como:

$$F_i = \frac{(\mu_{\alpha i} - \mu_{\beta i})^2}{\sigma_{\alpha i}^2 + \sigma_{\beta i}^2}$$

- F es mayor cuando los centroides de las clases $(\mu_\alpha$ y $\mu_\beta)$ están más separados (between classes).
- F es menor cuando la desviación de cada clase $(\sigma_\alpha$ y $\sigma_\beta)$ es mayor (within classes).

LDA biclase: Sobre las ideas del cociente de Fisher se puede buscar aquella dirección d sobre la que al proyectar todos los patrones x_1, x_2, \dots, x_n , se minimice el ratio de Fisher.

LDA multiclase: En el caso multiclase la solución consiste en encontrar la transformación d que hace máximo F_d . Se puede demostrar que esta transformación se obtiene con los autovectores con mayor autovalor del resultado de $S_w^{-1}S_b$, donde:

- S_b (between classes) es la matriz de covarianza de las medias de las clases respecto a la media de todos los patrones.
- S_w (within classes) es la suma de matrices de covarianzas para cada clase.

Algoritmo LDA:

- Entrada: Matriz X con n vectores de m características, vector de enteros y con las etiquetas de las c clases.
 - Salida: Matriz de proyección A
1. Calcular la matriz de dispersión intra-clase S_w
 2. Calcular la matriz de dispersión inter-clase S_b
 3. Calcular autovalores y autovectores de $S_w^{-1}S_b$

Devolver como matriz de proyección A la matriz con los autovectores con mayor autovalor. El número de autovectores devueltos será menor o igual a $\min(C-1, m)$

Redes neuronales

Definición de Neurona Artificial: Clasificadores que tratan de imitar el funcionamiento del cerebro. Una neurona artificial es una función que recibe una serie de valores de entrada, a los que aplica diferentes pesos, que utiliza para computar un valor de salida, mediante una función de activación.

Decisiones de diseño:

- Entrada (con un tamaño determinado por patrones a clasificar)
- Unidades ocultas (con tamaño determinado por la complejidad del problema a resolver): Número de capas y neuronas de cada capa.
- Salida (con tamaño determinado por el número de clases)
- Tipos de conexión entre capas
- Función de activación de cada neurona
- Función de salida de cada neurona
- Algoritmo de ajuste de pesos

Arquitecturas:

- Perceptrón multicapa
- Redes de Base Radial
- Redes de Hopfield
- Redes de Kohonen
- Time delay
- CNN
- LSTM

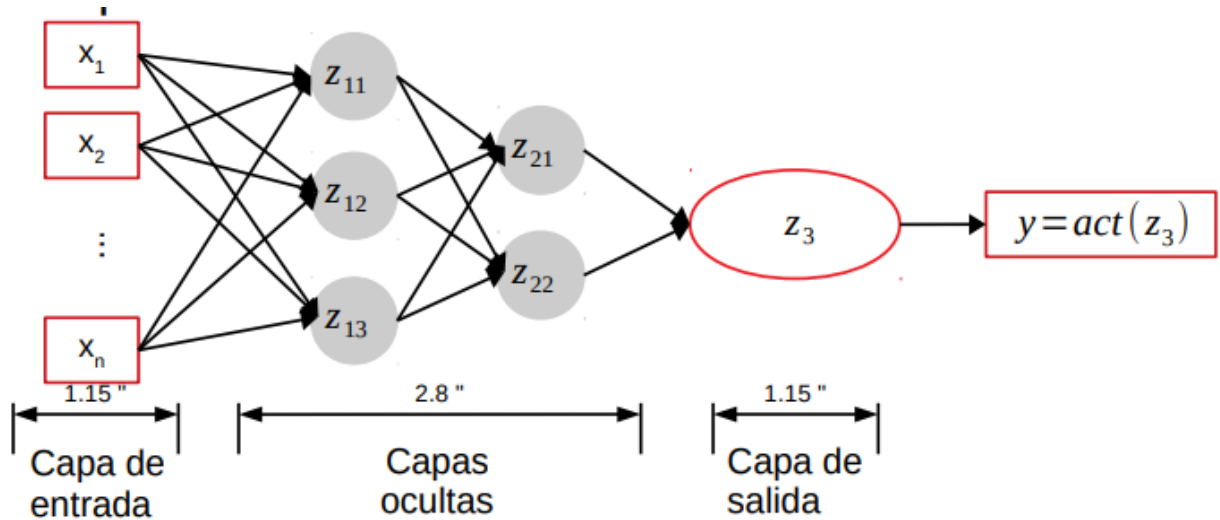
Perceptrón

Perceptrón de una capa: configuración que consta de una capa de perceptrones de salida únicamente conectadas a una capa de neuronas de entrada. El Perceptrón de una capa permite separación lineal entre clases (Se suele poner una neurona de salida por cada clase a clasificar y una neurona de entrada por cada característica discriminante)

Perceptrón multicapa: MLP, Capa Densa, Fully-Connected:

- Añadiendo más capas de neuronas ocultas a esta arquitecturas, se obtiene un perceptrón multicapa.
- organiza las neuronas en capas. En lugar de un único perceptrón, varias neuronas están conectadas con las mismas entradas x_1, \dots, x_n con diferentes pesos.

- Las salidas de esas neuronas son las entrada de una nueva capas de neuronas:



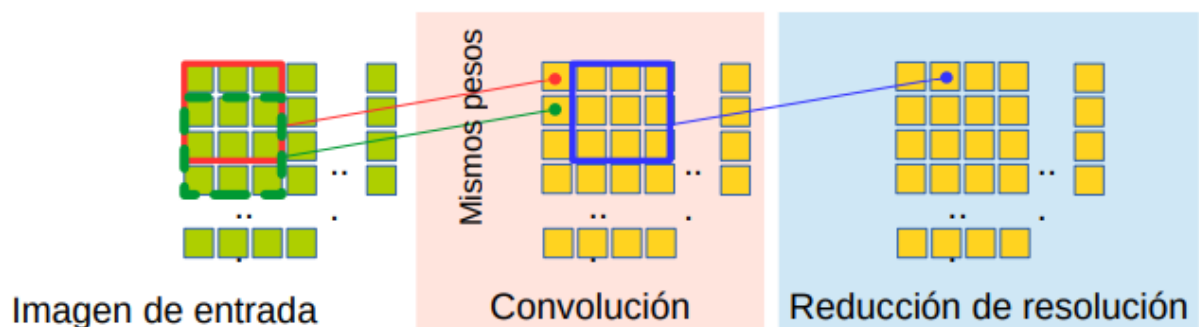
- Las características clásicas de la redes de neuronas son:
 - Robustez frente al ruido aleatorio (Selección automática de las características presentes a la entrada.)
 - Separación de regiones arbitrariamente complejas con una o más capas ocultas.
 - Capacidad de generalización (Buen interpolador)

Redes de convolución

En 1980 K. Fukushima introduce las redes de convolución (CNN) inspiradas en el estudio del campo visual humano. Las máscaras de convolución permiten abstraer características de las imágenes. Están formadas por capas de convolución, en las que se entrenan los pesos de las matrices de convolución.

Estructura de la CNN:

- La red de convolución se consigue forzando que determinados pesos adyacentes entre 2 capas tengan los mismos valores.
- Concatenan capas de convolución y de reducción de resolución. De esta forma tratan de abstraer características de la entrada.
- Cada capa abstrae una característica y por eso se llama mapas de características.



Capas convolucionales: parámetros:

- Núcleo (kernel): son las dimensiones de la máscara, filtro o matriz de convolución.
- Tamaño (size): : número de filtros de convolución

- Paso (stride): número de píxeles que se hace avanzar al filtro de convolución sobre la imagen
- Relleno (padding): manera de hacer que todos los píxeles de la imagen tengan la posibilidad de ser convolucionados, junto con su vecindad, con uno o varios filtro.

Tamaño imagen de salida: Combinando estos parámetros, se podrá controlar el tamaño de la imagen de salida a partir de la ecuación:

$$O = 1 + \frac{2P + I - K}{S}$$

- O : anchura (o altura) de la imagen de salida
- I : anchura (o altura) de la imagen de entrada
- K: tamaño del kernel
- P: Valor del relleno (1=padding same, 0=padding valid)
- S: stride

Capas de submuestreo: Con el fin de limitar el tamaño de la red e implementar invarianza ante distorsiones se realiza una operación de submuestreo (pooling). Algunas de las más comunes son:

- Max-Pooling
- Avg-Pooling

Arquitecturas basadas en CNN: LeNet

Aprendizaje profundo

Primeros éxitos de entrenamiento profundo:

- El término aprendizaje profundo (deep learning) aparece cuando Geoffrey Hinton y Ruslan Salakhutdinov mostraron cómo entrenar una red de neuronas con muchas capas.
- El idea fundamental tras el deep learning es el de representación distribuida: los conceptos a aprender se basan en otros conceptos más básicos que están en otro nivel de abstracción. Estos conceptos más básicos se basan a su vez en otros más básicos aún. Y así, sucesivamente.

Arquitecturas profundas:

- Deep neural networks.
- Convolutional neural networks.
- ResNet • Inception Networks
- Autoencoders
- LSTM
- Deep belief networks.

- Convolutional Deep Belief Networks.

Entrenamiento de la red: Para determinar los pesos de los perceptrones multicapa principalmente se utilizan algoritmos de descenso de gradiente:

- Backpropagation
- Backpropagation en batch
- Backpropagation con momentos
- RMSProp
- Adam

Estos algoritmos buscan minimizar el error que comete la red al clasificar un conjunto de patrones de entrenamiento

Batches: Para entrenar una red utilizando el algoritmo de retropropagación se puede operar de tres formas:

- Se podrían tomar todos los patrones de entrenamiento y calcular con todos ellos el mejor cambio de pesos para obtener la salida deseada.
- Se podría tomar un solo patrón, calcular el cambio de pesos para obtener la salida deseada para ese patrón y repetir este proceso sucesivamente con el resto de patrones (Stochastic Gradient Descent -SGD)
- Se podría tomar un subconjunto de los patrones (batch) y calcular con ellos el mejor cambio de pesos para obtener la salida deseada, luego se tomaría otro batch y así sucesivamente (Batch/Mini-Batch Gradient Descent).

Funciones de pérdida: Miden el error cometido durante el entrenamiento de la red.

- Problemas de regresión: Error Cuadrático Medio (MSE).
- Problemas de clasificación: Cross entropy.

Funciones de activación: Ajustan la salida al dominio del problema:

- Sigmoides: Variable continua de rango $[0,1]$.
- Tangente hiperbólica: variable continua de rango $[-1,1]$.
- Softmax: Variable continua que sea una distribución de probabilidad
- ReLU: evita problema desvanecimiento del gradiente. Adecuado para capas intermedias

Métodos de regularización: regularización es cualquier modificación que se hace a un algoritmo de aprendizaje que tiene la intención de reducir su error de generalización pero no su error de entrenamiento. Algunos de los más utilizados son:

- Aumentado de datos: Aumentar el conjunto de datos disponibles, añadiendo datos falsos creados a partir de los originales mediante transformaciones.

- Normalización del batch: Método de reparametrización adaptativa. Los modelos profundos se basan en la composición de varias funciones o capas. Se pueden obtener resultados inesperados en la actualización de las capas. Se resta la media del lote y se divide por su desviación típica. De este modo, se consigue que los parámetros se mantengan en un dominio adecuado para poder entrenar el modelo.
- Parada temprana : Se para el entrenamiento cuando se observa que el error de validación comience a ser mucho más grande que el de entrenamiento (sobreajuste)
- Dropout: La técnica de Dropout consiste en desactivar aleatoriamente algunas neuronas de la red durante el entrenamiento (previene el sobreentrenamiento). Hace que partes independientes de la red sean capaces de resolver el problema que se presenta, por lo que finalmente tenemos una mezcla de clasificadores. En redes de varias capas se suelen fijar diferentes porcentajes de dropout en cada capa. (A las capas cercanas a la salida se le pone un porcentaje mayor, y las cercanas a la entrada uno menor. Con esto forzamos a que las capas más profundas aprendan y por tanto también ayudan al problema del desvanecimiento del gradiente.)

Transfer learning y fine-tuning: La idea es empezar a entrar con los pesos pre-entrenados de arquitecturas conocidas con otras base de datos. Esto permite:

- Evitar definir la estructura de la red neuronal
- Tener que entrenarla desde cero
- Poder entrenar grandes arquitecturas con pocos datos.
- Selecciona una arquitectura neuronal que sabemos que es buena, y que ya ha sido entrenada. Una vez hemos seleccionado dicha arquitectura, quitamos las últimas capas y ponemos las nuestras, para adaptar la arquitectura a nuestro problema. El resto de parámetros permanecen “congelados”, no se entrenan. El fine-tuning, tiene un proceso similar, pero se descongela todo o parte del modelo.

Imagenet: En 2012 Krizhevsky supero la barrera del 25% sobre la base de datos de Imagenet utilizando backpropagation con algunos ajustes. Los ajustes que actualmente se consideran imprescindibles son:

- Uso de una función lineal rectificada (ReLU) en vez de la sigmoide o Tanh.
- Uso de Dropout.
- Uso de batches de patrones durante el entrenamiento (descenso del gradiente estocástico)
- Esta base de datos, al ser muy grande, es una de las más utilizadas para preentrenar arquitecturas y hacer transfer learning y fine-tuning