

Visión por computador

3 Trabajos de Teoría y Prácticas: 68 puntos

Motivación: 7 puntos

Proyecto final: 25 puntos

Introducción

La visión por computador comenzó como el análisis de imágenes, cuyo objetivo era aplicar una serie de operaciones para mejorar la imagen para que una persona pudiera verla mejor.

El objetivo de la VC es conseguir que una máquina llegue a analizar una imagen e interpretarla. Al igual que con el tratamiento de imágenes, con la VC interpretaremos el significado de una imagen a través de una matriz de datos que la representa.

Una imagen es una matriz de datos en la que podemos tener un solo dato o un conjunto de tres (RGB). Los algoritmos de interpretación nos permiten obtener algún tipo de conclusión observando los patrones que forman un conjunto de píxeles.

Historical and current approaches

1950 - 2010

Create good linear mathematical models

Most effort focused on good feature extraction

current

Learn from data the models

Internet is plenty of data to training models for many concepts

Local transformations

Realzar la imagen, encontrar los bordes, encontrar objetos...

Correlación convolución

Muestrear: extraer información hasta cierta frecuencia.

Como en aprendizaje automático, si tenemos una imagen con pocos píxeles no encontraremos un modelo que ajuste bien y en caso de tener una imagen con muchos el entrenamiento puede tardar un tiempo excesivo.

Escala: Obtenemos la información de un píxel junto con la de los vecinos.

Transformaciones locales:

$g(x,y) = g(x,y) + 20$ Hacemos la imagen más clara

$g(x,y) = g(-x,y)$ Darle la vuelta en el eje X

Filtrado de imágenes (Máscaras)

Noise reduction

El propio sensor genera ruido. En caso de tomar varias fotos podemos hacer el promedio del mismo píxel, por lo que la varianza tiende a 0 y es más seguro que tomar el valor promedio de los píxeles vecinos.

$$f(x,y) = \hat{f}(x,y) + g(x,y)$$

\hat{f} = imagen ideal

g = ruido

Primer intento

Suavizar la imagen, de esta forma rebajamos los picos que creemos que pueden ser ruido

Con o sin pesos.

Correlation mask

3x3 = 9 Box Filter

-1 0 1	-1 -1 -1
-1 0 1	0 0 0
-1 0 1	1 1 1

Uniforme - pesos equivalentes
 $1/(2k+1)^2$

H es una caja negra donde se realiza un procesamiento y obtenemos la información nueva.

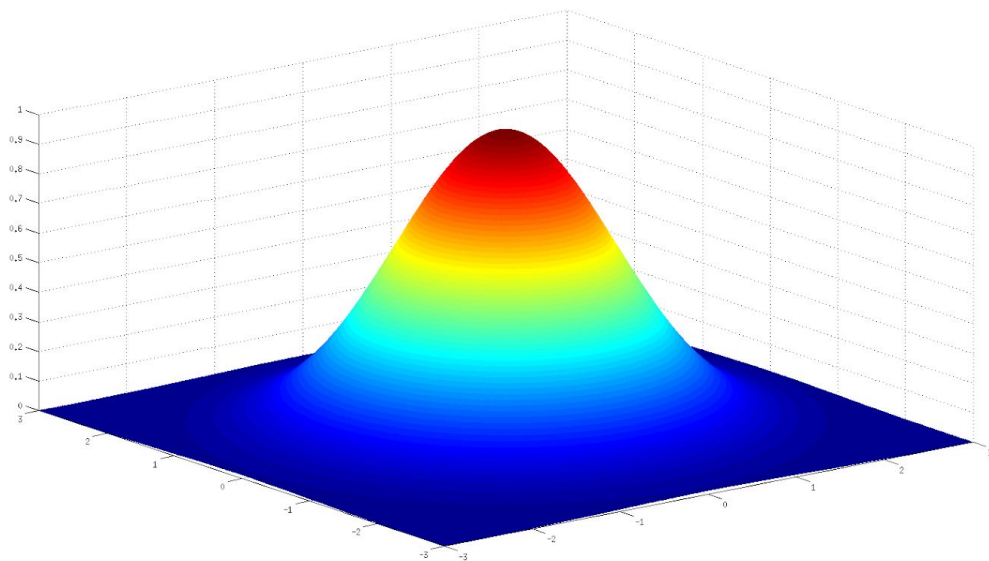
Cambiar los pesos a medida que pasamos el filtro o pasar muchas veces el filtro con distintos pesos

Cross-correlation

$$G = H \times F$$

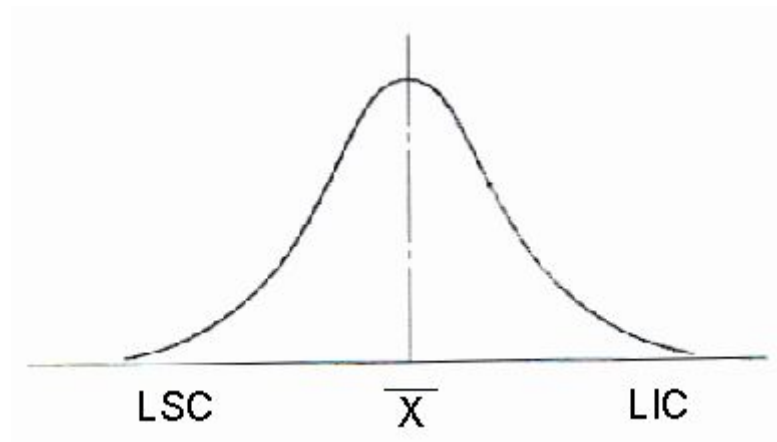
Los valores del kernel o la máscara deciden la información que obtenemos de la imagen.

Máscara gaussiana simétrica



El píxel central tiene más peso y los contiguos van decayendo conforme se alejan.

La gaussiana nunca alcanza cero, por lo que deberemos cortar la máscara para conseguir que los valores no cambien. Dependiendo de sigma variamos el tamaño de la máscara.



$$lsc = -k$$

$$lic = k$$

Buscamos un sigma que nos permita cortar nuestra máscara teniendo el máximo de información dentro.

$$k = 3$$

$$6 \text{ sigma} + 1$$

$$2 * (\text{entera } 3 \text{ sigma}) + 1$$

$$2k \text{ sigma} + 1$$

Tratamiento de bordes

Evitamos perder información al aplicar varios filtros.

clip filter, wrap around, copy edge, reflect across edge

Smooth = filtro paso bajo

Convolución

Es lo mismo que la correlación pero nos permite corregir la matriz que obtenemos al hacer una multiplicación matricial que invierte una matriz.

Gaussiana = isotrópica, da igual usar correlación o convolución

Correlación y convolución

Son filtros lineales que presentan las siguientes propiedades:

Shift invariant: El filtro no cambia los pesos al aplicarlo a la imagen

Superposition: Si tengo dos imágenes da igual pasarle los filtros a dos imágenes distintas y sumarlas que sumarlas y aplicar el filtro

Convolución:

Conmutativa, asociativa, distributiva, escalar, identidad

Correlación:

Cuando el patrón que define la máscara coincide en la imagen con una zona en la imagen que es muy parecida, la correlación es muy alta. Nos permite buscar modelos concretos de formas en la imagen.

No tiene asociativa.

cosas de furier

Convolución nos permite combinar filtros.

Correlación cruzada nos permite encontrar patrones o formas.

Señal

_____ ^ _____

 ^
 / \
____/ _____

Cuando la señal pequeña pasa por donde la señal grande tiene el pico más alto, ya que son las mismas señales.

Separabilidad

Máscara $M \times M$

Imagen $N \times N$

Coste de aplicar un filtro $O(M^2 N^2)$

Como la imagen no podemos tocarla, y debemos recorrer los N^2 píxeles para aplicar el filtro, vamos a buscar una forma de reducir la complejidad.

Recorriendo por filas y por columnas la máscara conseguimos que en cada píxel se realicen $2M$ operaciones, por tanto $O(M N^2)$

Matriz máscara = $\sum_{i=0}^{M-1} \sum_{j=0}^{M-1} u(i,j) v^t(i,j)$

En caso de matrices muy grandes la eficiencia puede ser menor que la transformada de Fourier discreta o dispersa, $O(N \log N)$

Smoothing filters

Median filter

Hybrid images

Dependiendo de la distancia podemos apreciar mejor las frecuencias altas o bajas.
Para hacer una imagen híbrida deberemos aplicar un filtro de paso bajo y un paso alto a dos imágenes y combinarlas.

Derivadas convolución

Epsilon distancia píxeles

Aproximar el valor de las derivadas a la función, no a las muestras.

$$f(x-1,y) - 2f(x,y) + f(x+1,y)$$

$$x-1 \quad x \quad x+1$$

$$(1, -2, 1)$$

Máscara sobre 3 píxeles

Máscaras de paso alto, suman 0.

Image gradient

Podemos calcular la dirección y magnitud de un punto en la imagen. Un vector es la derivada en x y en y de un punto.

Edge strength

Scale normalization

Conforme el sigma crece los valores de la gaussiana tiende a 0, ya que aplanamos la función.

Multiplicar por sigma en derivada primera.
 σ^2 en derivada segunda.

Template matching

Correlación cruzada da un valor que no es pequeño, aunque si el modelo no es exacto pixel a pixel tampoco será 1. No todas las fronteras de una imagen son las que delimitan los objetos, las sombras acechan.

Extracción fronteras -> filtro paso alto

Canny edge detector

Combinación umbral alto + bajo

Hysteresis

Large sigma detects large-scale edges, fronteras más representativas

Conforme el sigma aumenta, los hilos son más suaves.

Image representation.

Gaussian pyramid + laplacian pyramid

Upsampling

Coordenadas homogéneas: Completa el espacio incluyendo los puntos en el infinito (punto donde se juntan las vías del tren).

3D \rightarrow 2D

Transformaciones:

Traslación		
Escala		Rígidas, no pierden los ángulos
Semejanza		
Rotación		

Afines | Mantiene el paralelismo

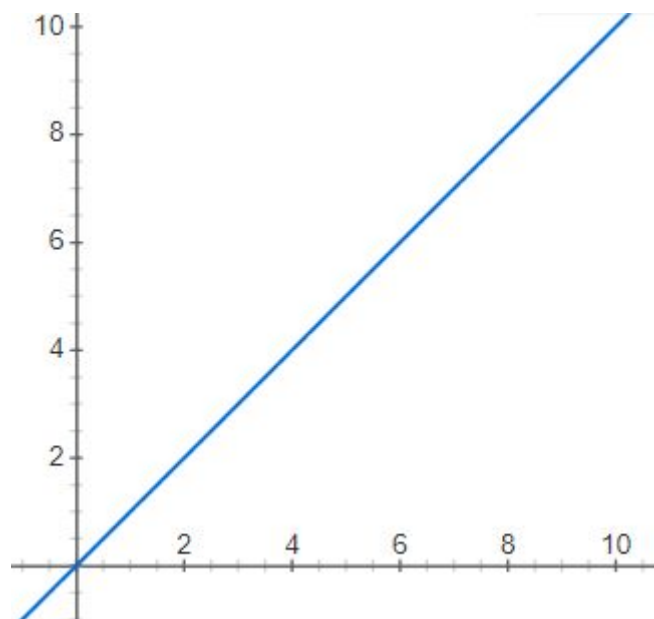
Espacio 2D proyectivo = 2D + coordenada homogénea

$(x,y,w) = (x/w, y/w)$ para $w \neq 0$

El conjunto de las rectas que pasan por un punto $(k_x, k_y, 1)$, define los puntos en el infinito.

$(x,y,w) = (x/w, y/w, w/w) \rightarrow (k_x, k_y, 1) = (x,y,1/k)$

Recta que llega al mismo punto en el primer plano y pasa por los puntos (k_x, k_y)



Cuando un punto está en el infinito $(1/k)$ vale 0

$$ax + by + c = 0 \rightarrow (a,b,c)(x,y,1) = 0$$

Encontrar un vector (a,b,c) que al multiplicarlo por un punto sea 0. Recta en el infinito.
Recta (0,0,1).

$$(0,0,1)(x,y,0) = 0$$

a b c
d e f Proyectividad, deforma los puntos en el infinito
g h i

3x3 con determinante $\neq 0$

Si dos rectas se cortan en el espacio original también se cortan en el proyectivo. Una recta siempre será recta, no se puede curvar.

a b c
d e f = H $\det(H) \neq 0$
g h i

$$\begin{pmatrix} x' \\ y' \\ w' \end{pmatrix} = H \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

Cómo hallar H (?)

Diferentes escenas cambia punto de vista - transformaciones afines para trabajar con los puntos de vista

Homografía: H (3x3) $\det H \neq 0$

$x^T \cdot l$: El producto escalar de dos vectores de 3 dimensiones es la ecuación de una recta

$$x = (x,y,1), l = (a, b, c)$$

$$x \rightarrow \{k(x, y, 1), k \neq 0\}$$

$$x^T I = 0 \quad (x, y, 1/k), \quad k \rightarrow \infty$$

$$\text{Dado } x' = Hx$$

$$\text{lambd} \cdot \begin{matrix} x' \\ y' \\ 1 \end{matrix} = H \begin{matrix} x \\ y \\ 1 \end{matrix}$$

$$A = \begin{bmatrix} x & y & 1 & 0 & 0 & 0 & -xx' & -yx' & -x \\ 0 & 0 & 0 & x & y & 1 & -xy' & -yy' & -y \end{bmatrix} \quad h = [a \ b \ c \ \dots \ h \ i]$$

$$\begin{matrix} A & h & = 0 \\ 8 \times 9 & 9 \times 1 \end{matrix}$$

Una matriz multiplicada por una constante de proporcionalidad (lambda), aporta la misma transformación. Existen infinitas matrices que realizan la misma transformación y todas dependen de la constante de proporcionalidad.

Si tenemos más de 4 puntos para resolver la ecuación y tenemos mismo rango que número de incógnitas no podemos resolverlo de forma exacta.

Transformamos el problema a uno de aproximación.

$$\begin{matrix} A & h = \text{epsilon} \text{ (vector de valores cercanos a cero)} \\ 2n \times 9 & 9 \times 1 \end{matrix}$$

$$A h = \text{eps} \Rightarrow \text{Minimizar } \|Ah\|$$

$$\text{norma de } h \neq 0 \Rightarrow \|h\| \neq 0$$

Como todas las matrices son iguales, en vez de comparar que sea distinto de 0 vamos a comprobar si la norma de h es 1, por lo que demostramos que no todos los valores son 0.

Mínimos cuadrados

$$\begin{matrix} \text{Min } \|Ah\| \\ \|h\|=1 \end{matrix}$$

Transformación en valores singulares

Descomposición en valores singulares de A - Equivalente a mínimos cuadrados

$$\text{SVD}(A) = UDV^t$$

Columna V asociada a mínimo de D

Comparar 2 transformaciones - Elegir la misma constante de proporción y comparar si dan los mismos valores.

DLT: Hartley Arzisen

Ruido gaussiano: Ruido que introducimos al realizar una medida

$\{(x_i, y_i), i = 1 \dots n \text{ en correspondencia}\}$

Buscamos minimizar el ruido gaussiano

Si en una matriz tenemos valores constantes, el sistema va a ser determinado y con solución única.

Image alignment algorithm

Valorar cómo los valores erróneos influyen - outliers

mins cuadrados suma de errores, minimizar los errores más grande para que la suma sea menor

Rectas que pasen por todos los puntos.

Rectas más robustas aquellas cuya distancia con los puntos sea mínima - ran sac

Ran Sac

Fijar homografía

Medir errores respecto del modelo inliers

Cuantos modelos

Pros

simple y general

aplicable a muchos problemas

trabaja bien con la practica

implementaciones en tpo real

Buscamos que haya muchos puntos iguales en ambas imágenes, esto se puede deber a que sean imágenes tomadas desde distintos puntos de vista....

Harris detector, en una imagen hay una región en la que encontramos dos gradientes bastante fuertes

Matriz Harris: Matriz primeras derivadas

Harris funciona bien con : giros
 cambio intensidad luminosa, $I = aI + b$

Criterio Harris en problemas de escala

Si hacemos zoom a una esquina, la esquina desaparece y tenemos pequeños trozos que serán clasificados como bordes.

Corner = 2 gradientes

Scale invariant detection

Cogemos distintos tamaños de escala (5x5, 9x9,...)

Para calcular las esquinas calculamos la matriz de derivadas, que es la suma de la derivada de todos los puntos. Si tenemos una escala muy grande y una esquina pequeña, no la detectará.

Automatic scale selection

Elegir una función con un máximo. Conforme cambiamos la escala el algoritmo será más o menos relevante, elegiremos aquel que sea más relevante.

Escoger aquella escala en la que tengamos el mayor valor de harris.

Para la misma imagen con zoom, en el mismo punto que antes tendremos la misma escena (la región que teníamos antes ampliada por el mismo factor que la imagen) pero con más información que antes.

Regiones en correspondencia

Normalize: rescale to fixed size

Implementaciones

1. Fijo el tamaño de la ventana y voy reduciendo la imagen - pirámide gaussiana
Reducir la imagen 3 o 4 veces, la imagen debe tener un tamaño mínimo.
2. Aumentar el tamaño de la ventana

Blob detection in 2D

Laplacian of Gaussian: Circularly symmetric operator for blob detection in 2D
Corregir con escala

Escoger rango de valores de sigma para los que aplicar la ventana de Harris

Buscar puntos que sean máximos respecto la escala anterior o la siguiente, encontrar el máximo de los cuadrados de la laplaciana.

Podamos puntos que no tengan un máximo que sea fácil de encontrar (funciones estrictamente crecientes) o que no tengan varios máximos puesto que no sabríamos cual elegir, de esta forma elegiremos un conjunto pequeño de puntos con los que trabajar.

Efficient implementation

Laplaciana en función del calor

Laplacian vs Difference of Gaussians

Lowe's pyramid scheme
3 o 4 niveles de Gaussiana

Al hacer varios niveles buscamos el máximo de todas las capas (Cubo 3D con los valores de cada escala).

Affine adaptation

Qué pasa si la deformación entre las imágenes es fuerte - usar elipses

Con círculos buscamos el máximo

Con elipses buscamos las regiones con mayor gradiente, cosa que podemos volver a encontrar si cambiamos la imagen

Regions Detection MSER

Feature descriptors

We know how to detect good points but... how to match them

Scale Invariant Feature Transform - SIFT

Que el descriptor esté pegado a los gradientes

Todos Empíricos, cogemos el que mejor lo hace

Normalizar lo que extraes para poder comparar imágenes

Feature matching
(Depende del volumen de puntos escogemos un algoritmo u otro)

Distancia - diferencia entre dos características

Mediante la norma cuadrática de la distancia de las características buscamos aproximar el umbral al máximo para asegurar que es la misma característica

Al analizar localmente perdemos la información que nos da la escala, por lo que en imágenes con un patrón repetitivo podemos encontrar varias regiones que sean iguales.

kdtree no garantiza el punto más cercano

Juntar panorámica

Detectar características

Combinar características de imágenes

Homografía

Combinar imágenes

Reconocimiento de Objetos

Objetos concretos - instancias

- Detección de regiones
- Extraemos información de la imagen en términos de los descriptores.
- Modelo = Bolsa de palabras + {
if - tdf
Geometría
}

Query expansion - A partir de una imagen explotar la base de datos para encontrar imágenes similares.

Recognition via alignment

Pros: Efectivos y simples (poca complejidad). Es un modelo corto que no necesita mucho tiempo.

Cons: Scaling with number of models. Spatial verification as post-processing - not seamless, expensive for large-scale problems. Not suited for category recognition.

tf-idf weighting

Object categorization

Segmentar - Qué es un objeto y dónde está

Object recognition is it really so hard? - Falsos positivos

Ortogonal - Todo lo que aporta es distinto
Independiente - Aporta nueva información