

HOG

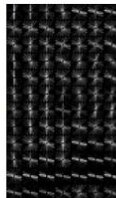
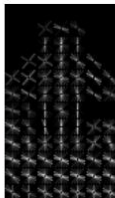
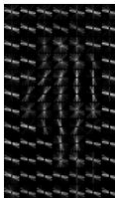
Elementos de Reconocimiento Visual

Departamento de Computación

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

2do Cuatrimestre 2023



Reconocimiento de Personas según Dalal & Triggs 2005



Histogram of Oriented Gradients

ref

Histograms of Oriented Gradients for Human Detection

Navneet Dalal and Bill Triggs

<http://lear.inrialpes.fr>

Qué es un Descriptor de Características?

Un descriptor de características es una representación de una imagen o de una porción de imagen que simplifica la imagen extrayendo información útil y desechando información superflua.

Cómo funciona?

HOG es un descriptor de características que convierte una imagen de tamaño ancho x alto x 3 (canales) en un vector / matriz de características de longitud n . Para HOG, la imagen de entrada tiene un tamaño de $64 \times 128 \times 3$ y el vector de características de salida tiene una longitud de 3780.

Qué es un Descriptor de Características?

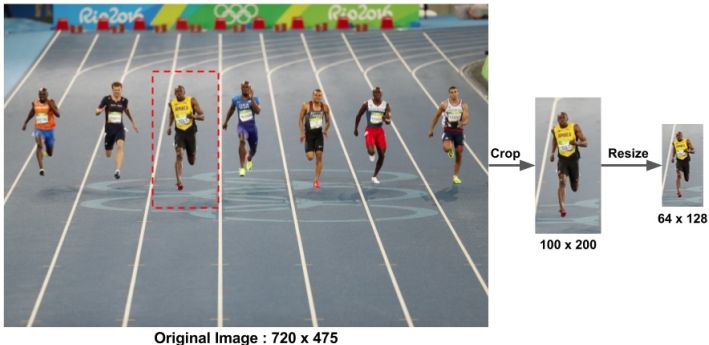
Cómo funciona?

En HOG, la distribución (histogramas) de direcciones de gradientes (gradientes orientados) se utilizan como características. Los gradientes (derivadas x e y) de una imagen son útiles porque la magnitud de los gradientes es grande alrededor de los bordes y esquinas (regiones de cambios abruptos de intensidad) y es allí adónde hay más información, lo contrario que las regiones planas.

Calculando el descriptor HOG

1. Preprocesamiento

A partir de una imagen grande, se selecciona un patch de tamaño 100×200 para calcular el descriptor de funciones HOG. El patch se recorta de una imagen y se redimensiona a 64×128 .



Calculando el descriptor HOG

2. Cálculo de Gradiente



Izq: Valor absoluto del gradiente respecto de x . Centro: Valor absoluto del gradiente respecto de y . Der: Magnitud del gradiente.

- Los gradientes destacan las líneas verticales y horizontales respectivamente según $\frac{df}{dx}$ y $\frac{df}{dy}$,
- La magnitud destaca donde hubo cambios de intensidad sin importar la dirección.
- Si la región es lisa no hay orientación, no hay información relevante. Por lo tanto la imagen gradiente eliminó información no esencial (por ejemplo, fondo de color constante), pero resaltó los contornos.
- En cada píxel, el gradiente tiene una magnitud y una dirección. Para las imágenes en color, se evalúan los gradientes de los tres canales (como se vio en la figura anterior).
- La magnitud del gradiente en un píxel es el máximo de la magnitud de los gradientes de los tres canales y el ángulo es el ángulo correspondiente al gradiente máximo.

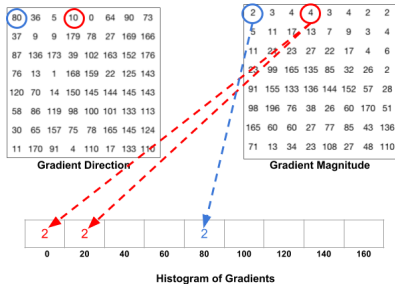
Calculando el descriptor HOG

3. HOG en celdas de 8x8

La imagen se divide en celdas de 8×8 y se calcula el histograma de gradientes para cada celda.



Un patch de imagen de 8×8 contiene valores de $8 \times 8 \times 3 = 192$ píxeles. El gradiente de este patch contiene 2 valores (magnitud y dirección) por píxel, lo que suma $8 \times 8 \times 2 = 128$ números.



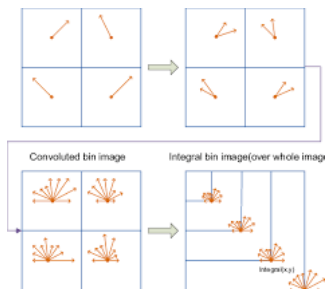
El histograma contiene 9 bins correspondientes a los ángulos 0, 20, 40... 160.

La dirección selecciona un bin y el voto (el valor que acumula el bin) se selecciona en función de la magnitud. Ej: píxel rodeado de azul: ángulo de 80 grados y magnitud de 2, entonces suma 2 al bin 80; píxel rodeado de rojo, ángulo de 10 grados y magnitud de 4. Allí aplica una interpolación 3-D

La interpolación utilizada por Dalal consiste en considerar que los bins tienen ancho b , llama w a la magnitud, x_1 y x_2 son bins vecinos, consecutivos, y x es la orientación, tal que $x_1 \leq x < x_2$, cada w lo distribuye según la fórmula:

$$\mathbf{h}(x_1) \leftarrow \mathbf{h}(x_1) + w \cdot \left(1 - \frac{x - x_1}{b}\right)$$

$$\mathbf{h}(x_2) \leftarrow \mathbf{h}(x_2) + w \cdot \left(\frac{x - x_1}{b}\right)$$



80	36	5	10	0	64	90	73
37	9	9	179	78	27	169	166
87	136	173	39	102	163	152	176
76	13	1	168	159	22	125	143
120	70	14	150	145	144	145	143
58	86	119	98	100	101	133	113
30	65	157	75	78	165	145	124
11	170	91	4	110	17	133	110

Gradient Direction

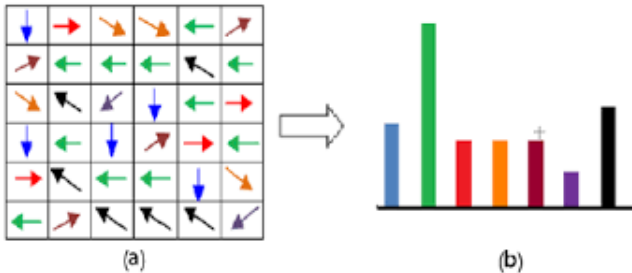
2	3	4	4	3	4	2	2
5	11	17	13	7	9	3	4
11	21	23	27	22	17	4	6
23	99	165	135	85	32	26	2
91	155	133	136	144	152	57	28
98	196	76	38	26	60	170	51
165	60	60	27	77	85	43	136
71	13	34	23	108	27	48	110

Gradient Magnitude



Histogram of Gradients

Entre todos los pixels de la celda de 8x8 contribuyen al histograma final.



Calculando el descriptor HOG

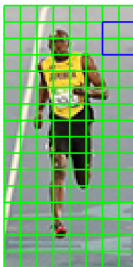
4. 16×16 Block Normalization

Normalización

- HoG está normalizado en contraste
- aplicar la norma al vector de features da mayor precisión en iluminación e intensidad.
- Hog es invariante a transformaciones geométricas y fotométricas, excepto a la orientación del objeto. Muy adecuado para la detección humana en imágenes.

Pasos

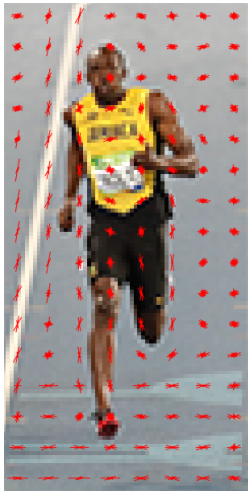
- Para normalizar utilizaron distintas normas: L2-norm , L2 Hyst
- Si bien podemos normalizar el histograma de 9×1 , utilizaremos bloques de 16×16
- los bloques de 16×16 tienen 4 histogramas de 9×1
- concatenando los 4 histogramas se obtiene un vector de 36×1
- y es este vector el que se normaliza



5. El vector característico de HOG

Para calcular el vector de características final para todo el patch de imagen (el de 128×64), los vectores de 36×1 se concatenan en un vector gigante. ¿Cuál es el tamaño de este vector?

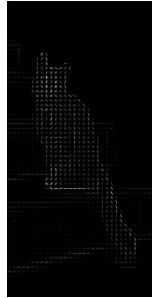
- Hay 7 posiciones horizontales y 15 verticales, lo que hace un total de $7 \times 15 = 105$ posiciones.
- Cada bloque de 16×16 está representado por un vector de 36×1 . Entonces, cuando los concatenamos todos en un vector de dimensión $36 \times 105 = 3780$.



HOG

Elementos de Reconocimiento Visual, DC-FCEyN-UBA

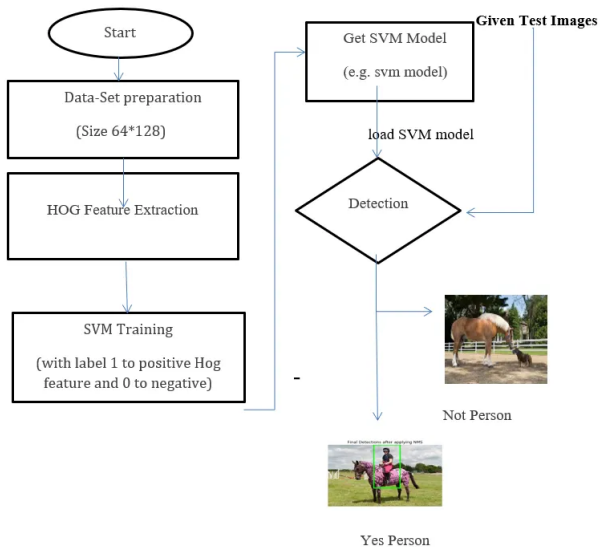
Ejemplo: Imagen, redimensionar y mostrar la
matriz de orientaciones



- HOG es muy usado como descriptor de objetos en imágenes. Puede detectar caras, automóviles, matrículas, señales de tráfico, pájaros, etc.
- El descriptor HOG, en el ejemplo un vector 3780 posiciones, es el vector de características en cualquier esquema de clasificación de aprendizaje automático.

Algunas contras de HoG

- Son necesarios muchos ejemplos, eso significa etiquetar muchas imágenes.
- En la práctica, es un poco más difícil utilizar un detector basado en HOG, ya que en realidad rara vez se da el caso de que la posición y el tamaño del objeto en la imagen se conozcan a priori.
- Hay que escanear la imagen completa en todos los tamaños posibles y para cada subimagen clasificar el HOG resultante.
- Como no es una tarea fácil, por eso se proveen de datasets etiquetados positivos y negativos diseñados para reconocimiento.



La tarea de este trabajo será implementar un reconocedor de personas basados en el artículo de Dalal - Triggs

N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

Link para bajarse las imágenes:

https://drive.google.com/file/d/14GD_pBpBsprPiZlkmtXN_y5K72To16if/view?usp=sharing