

# HOG

2018/9/2

JunYu Wang

## 1. HOG

### Histogram of Oriented Gradients

计算出每一个像素的梯度大小和梯度方向，主要刻画出图像的 lines & corners 等 useful 的信息，并且忽略颜色等 extraneous 的信息

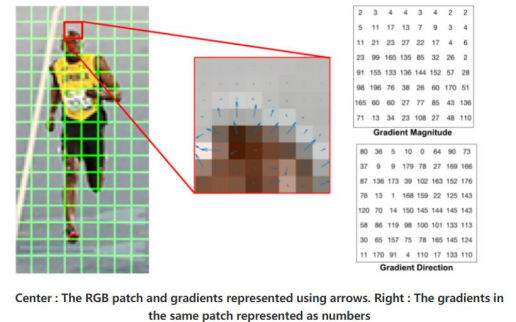
## 2. 计算 HOG

**Step1.** At every pixel, the gradient has a magnitude and a direction. For color images, the gradients of the three channels are evaluated ( as shown in the figure above ). The magnitude of gradient at a pixel is the maximum of the magnitude of gradients of the three channels, and the angle is the angle corresponding to the maximum gradient.

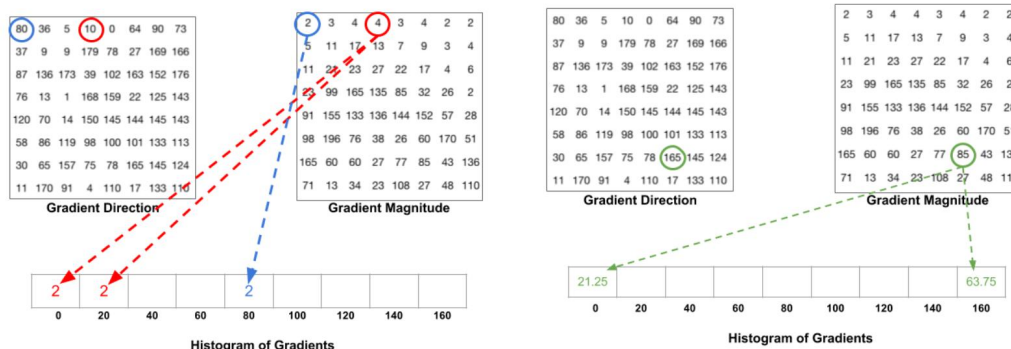
$$\textcircled{1} \quad g = \sqrt{g_x^2 + g_y^2} \quad \text{Gradient Magnitude}$$

$$\textcircled{2} \quad \theta = \arctan \frac{g_y}{g_x} \quad \text{Gradient Direction}$$

$\textcircled{3}$  Using  $[-1,0,1]$  &  $[1,0,1]^T$  kernel to calculate  $g_x$  &  $g_y$



**Step2.** In this step, the image is divided into  $8 \times 8$  cells and a histogram of gradients is calculated for each  $8 \times 8$  cells. An  $8 \times 8$  image patch contains  $8 \times 8 \times 3 = 192$  pixel values. The *gradient of this patch contains 2 values* ( magnitude and direction ) per pixel which adds up to  $8 \times 8 \times 2 = 128$  numbers. By the end of this section we will see how these *128 numbers are represented using a 9-bin histogram* which can be stored as an array of 9 numbers. (The number of bins is chosen by ourselves )



**Step3. Normalization** 由于光线等等的影响，magnitude 可能会有不同。因此我们使用

normalization to “normalize” the histogram so they are not affected by lighting variations.

前面提到，我们把图片切成  $8 \times 8$  的像素 block。接着我们用  $16 \times 16$  的 kernel 滑动。因此原本的 9 个 bin 的向量就会变成 36 长度的向量，并被我们 normalize

E. g.  $[128, 64, 32]$   $\sqrt{128^2 + 64^2 + 32^2} = 164.64$  正则化处理后得到  $[0.87, 0.43, 0.22]$

1. How many positions of the  $16 \times 16$  blocks do we have ? There are 7 horizontal and 15 vertical positions making a total of  $7 \times 15 = 105$  positions.

2. Each  $16 \times 16$  block is represented by a  $36 \times 1$  vector. So when we concatenate them all into one gaint vector we obtain a  $36 \times 105 = 3780$  dimensional vector.

这样就得到了我们的 HOG feature

