

图像理解中的模式识别方法 - 方向梯度直方图（HOG）

谢博琛

参考资料:

- 1 参考书《图像处理、分析与机器视觉（4th）》P367 – P369
- 2 Learn OpenCV - <https://www.learnopencv.com/histogram-of-oriented-gradients/>
- 3 Histograms of oriented gradients for human detection, CVPR 2005
- 4 CSDN - HOG 特征（Histogram of Gradient）学习总结 - <https://blog.csdn.net/u011285477/article/details/50974230>
- 5 简书 - 方向梯度直方图（HOG） - <https://www.jianshu.com/p/6f69c751e9e7>

Part 1: What's the HOG?

HOG = Histogram of Oriented Gradients 方向梯度直方图

HOG 是一种图像特征描述子。

首次提出：Navneet Dalal & Bill Triggs, CVPR 2005, Histograms of oriented gradients for human detection。

作用：应用在计算机视觉和图像处理领域，用于目标检测的特征描述器。

Part 2: Object Detection and Location Based on HOG

【全局流程】



1. 确定窗口，胞体，块大小/形状和重叠大小

尺寸关系：patch（窗口） > block（块） > cell（胞体） > pixel（像素）。

扫描时候用块 block，统计特征时候用 cell。

需要注意的是，特征是通过互相重叠的块计算出来的，所以必须设计一个网格来指定重叠的参数。

确定窗口，胞体，块大小/形状和重叠大小。根据目标检测的任务不同，确定图像窗口的大小和形状（如图 10.31 所示，行人检测通常用 64×128 的窗口；窗口与感兴趣的物体之间应该有足够大的间隔-行人检测中 16 个像素大小的间隔是合适的）。

胞体中相邻的像素包含了图像的局部信息，所以必须确定胞体的大小和形状。胞体一般包含 6×6 到 8×8 个像素（人体关节约为 6 到 8 个像素宽），并且在行人检测中一个块中一般包含 2×2 或 3×3 个胞体。另外，块（胞体）可以是方形或者圆形的。图 10.30 展示了方形和圆形的块结构。方形的胞体比较常用因为他们计算效率高。另外，特征是通过互相重叠的块计算出来的，所以必须设计一个网格来指定重叠的参数。

2. 光度规则化（图像归一化处理）

本方法采用了对比度归一化来降低对光照，阴影或其他光度转换的敏感性。

光度规则化。使用全局的图像数据归一化或者伽玛校正来处理整个图像。合适的话推荐使用彩色图像（多频带）并对每一个通道独立地进行伽马校正。

为了提高检测器对光照等干扰因素的鲁棒性，需要对图像进行 Gamma 校正，以完成对整个图像的归一化，目的是调节图像的对比度，降低局部光照和阴影所造成的影响，同时也可以降低噪音的干扰（当 r 取 $1/2$ 时，像素的取值范围就从 $0\sim255$ 变换到 $0\sim15.97$ ）。

Gamma 校正的公式如下所示：

$$I(x, y) = I_0(x, y)^{1/r}$$

3. 计算方向梯度直方图和形成块 HOG 描述子

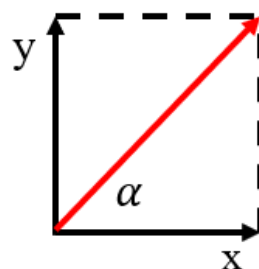
Step 1: Gradient Computation 梯度计算

首先，进行梯度计算，输入图像在像素点 (x, y) 的水平方向 x 和竖直方向 y 的梯度计算公式为：

$$G_x(x, y) = I(x + 1, y) - I(x - 1, y)$$

$$G_y(x, y) = I(x, y + 1) - I(x, y - 1)$$

像素点 (x, y) 处的梯度幅值和方向为：



$$\nabla G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}$$

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right)$$

一般都会用水平方向 $[-1, 0, 1]$ 和竖直方向 $[1, 0, -1]^T$ 边缘检测算子对输入图像进行卷积运算，从而得到水平方向和竖直方向上的梯度分量，算出每个边缘像素点处的梯度幅值和方向。

在每个像素点，都有一个幅值（magnitude）和方向，对于有颜色的图片，会

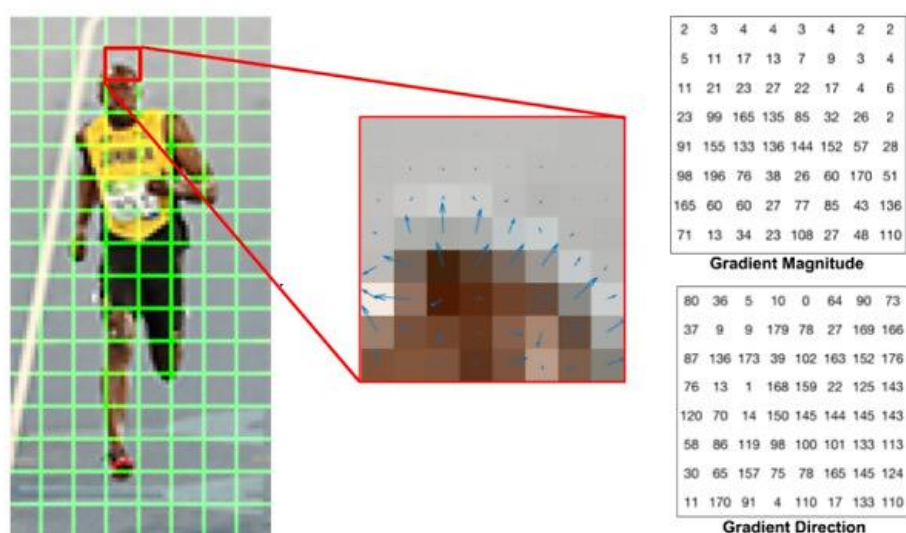
在三个 channel 上都计算梯度。那么相应的幅值就是三个 channel 上最大的幅值，角度（方向）是最大幅值所对应的角。

Step 2: 在胞体 cell 上计算方向梯度直方图

这一步骤的目的是为局部图像区域提供一个指示函数量化梯度方向的同时能够保持对图像中人体对象的姿势和外观的弱敏感性。

一般情况下，HOG 特征用于行人检测，窗口 patch 一般选择 64×128 个像素，每 8×8 个像素组成一个胞体 cell，每 2×2 个胞体 cell 组成一个块 block，那么扫描窗口的步长则为 8 个像素（以胞体 cell 为最小单元）。所以，扫描窗口在水平方向需要扫描 7 次，在竖直方向需要扫描 15 次，在整个图像需要扫描 $7 \times 15 = 105$ 次。

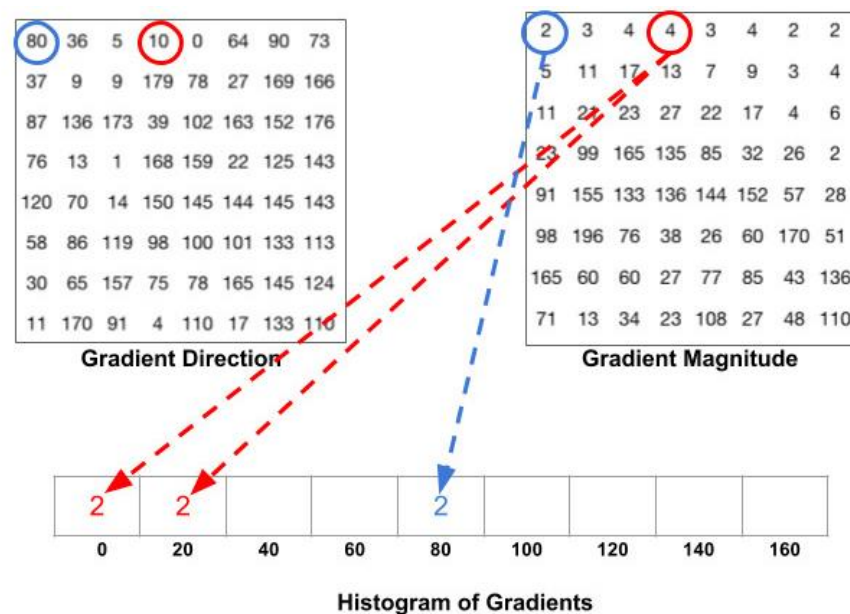
单独将其中一个最小单元胞体 cell (8×8 个像素) 的小格拿出来，定义方向梯度中的方向范围为 $0 - 180^\circ$ （为无符号梯度， $0 - 360^\circ$ 则为有符号梯度），为了画方向梯度直方图，我们还需要选取合适的组距也就是 bin，这里组距选取 $\pi/9$ ，也就是最后的直方图组数 bin 为 9。我们将像素的梯度幅值作为权值进行投影，用梯度方向决定向哪一维进行投影。



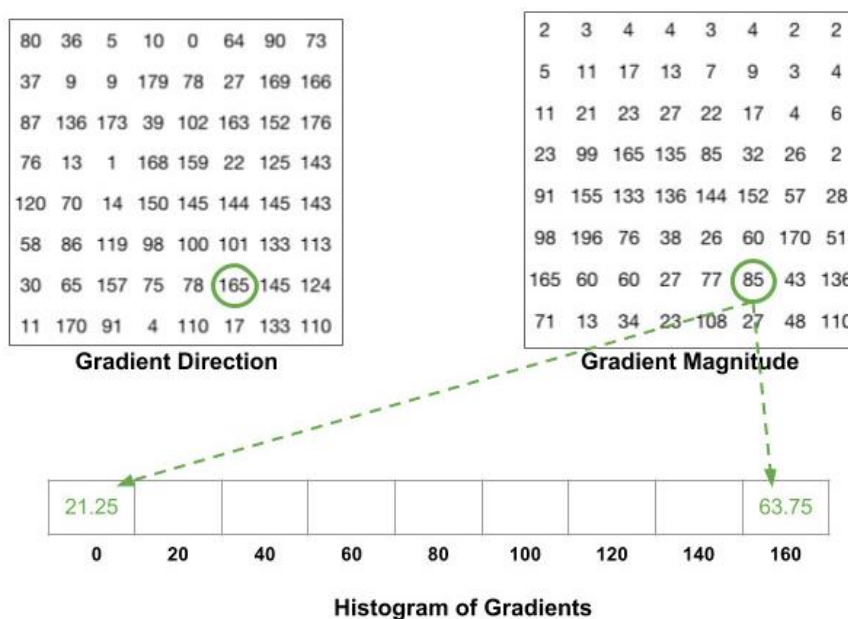
Center : The RGB patch and gradients represented using arrows. Right : The gradients in the same patch represented as numbers

下图是一个细胞单元内的方向梯度直方图，角度分辨率是在 180° 的范围内，以 20° 等分，即一个细胞单元的 HOG 特征是一个 9 维的向量，9 个 bin 代表的角度是 $0^\circ, 20^\circ, 40^\circ, 60^\circ, 80^\circ, 100^\circ, 120^\circ, 140^\circ, 160^\circ$ 。此前，我一直怀疑明

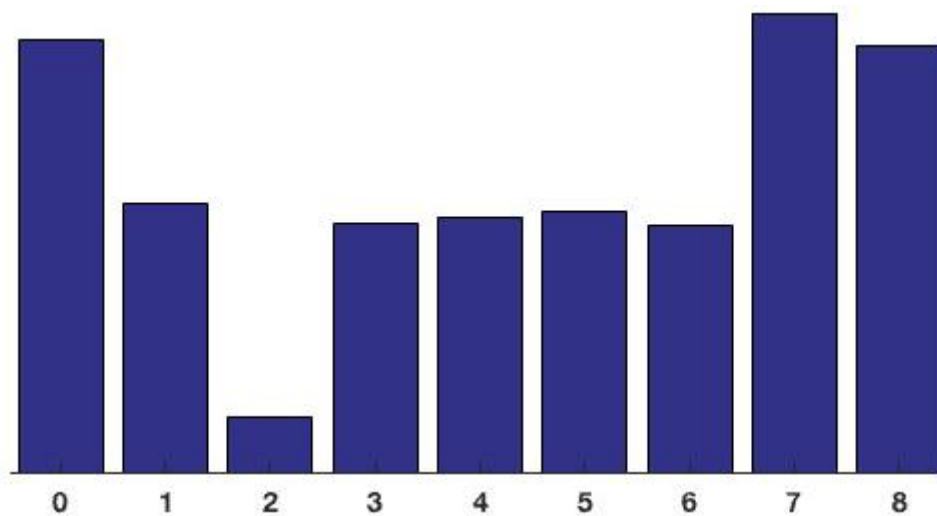
先来看蓝色圆圈圈出来的像素点，它的角度是 80，幅值是 2，所以它在第五个 bin 里面加了 2。再来看红色圈出来的像素点，它的角度是 10，幅值是 4，因为角度 10 介于 0-20 的中间(正好一半)，所以把幅值一分为二地放到 0 和 20 两个 bin 里面去。



当梯度方向大于 160° 时，就是我们刚才提到的特殊情况。



把一个 cell 中 $8*8=64$ 个像素的梯度幅值和方向，表示在 9 bin 的直方图中，如图所示：



Step 3: 块 block ($16*16$ pixels = $2*2$ cells) 的归一化

理想情况下，我们希望描述符独立于光照变化。换句话说，我们想对方向梯度直方图，以使它们不受光照变化的影响。

Let's say we have an RGB color vector $[128, 64, 32]$. The length of this vector is $\sqrt{128^2 + 64^2 + 32^2} = 146.64$. This is also called the L2 norm of the vector.

Dividing each element of this vector by 146.64 gives us a normalized vector $[0.87, 0.43, 0.22]$. Now consider another vector in which the elements are twice the value of the first vector $2 \times [128, 64, 32] = [256, 128, 64]$. You can work it out yourself to see that normalizing $[256, 128, 64]$ will result in $[0.87, 0.43, 0.22]$, which is the same as the normalized version of the original RGB vector. You can see that normalizing a vector removes the scale.

既然我们知道如何对向量进行归一化，您可能会想起，在计算 HOG 时，您可以像对上述 $3*1$ 向量进行归一化一样简单地对 $9*1$ 直方图进行归一化。这不是一个坏主意，但是更好的主意是对更大的 $16*16$ 块进行标准化。 $16*16$ 块具有 4 个直方图，可以将它们连接起来以形成 $36*1$ 元素向量，并且可以像对 $3*1$ 向量进行归一化的方式对其进行归一化。

Step 4: 形成块 HOG 特征描述子

在 Step 2 中我们已经提到过，一般情况下，HOG 特征用于行人检测，窗口 patch 一般选择 $64*128$ 个像素，每 $8*8$ 个像素组成一个胞体 cell，每 $2*2$ 个胞体 cell 组成一个块 block，那么扫描窗口的步长则为 8 个像素（以胞体 cell 为最小

单元)。所以，扫描窗口在水平方向需要扫描 7 次，在竖直方向需要扫描 15 次，在整个图像需要扫描 $7 \times 15 = 105$ 次。

所以，我们有 $7 \times 15 = 105$ 个块 block，每个块 block 都有 36×1 的特征向量，所以，一个 64×128 的窗口 patch 含有了 $7 \times 15 \times 36 = 3780$ 维特征向量，也可以称为块 HOG 特征描述子。

4. 分类

分类：使用任何基于特征的分类器都可以用 HOG 描述子进行训练和识别，HOG 描述子在行人检测/定位中使用高效的线性分类器-线性支持向量机效果很好[Dalal and Triggs, 2005]，见图 10.32 (f) 和图 10.32 (g)。

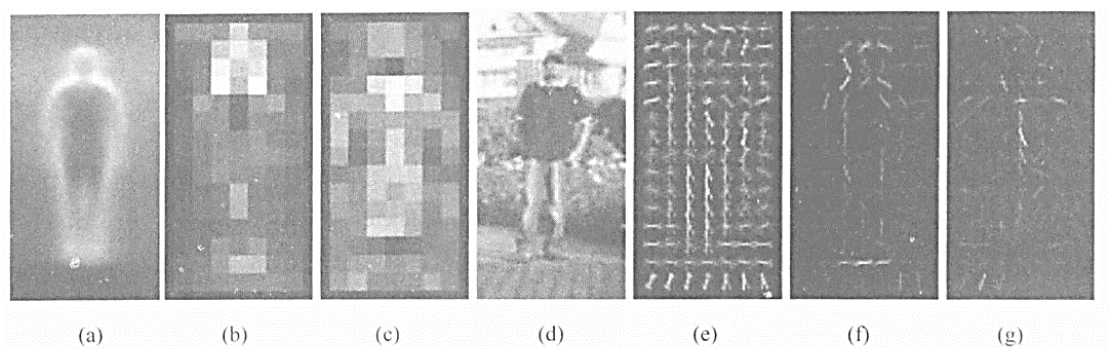
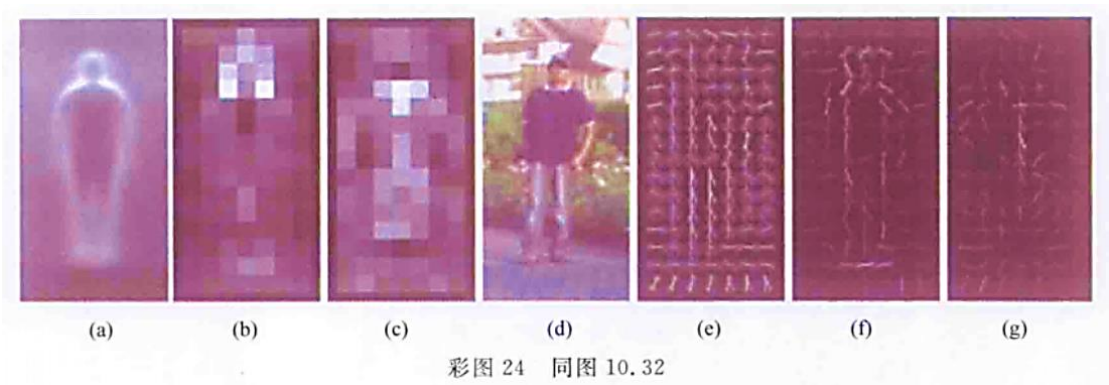


图 10.32 行人检测中的 HOG 描述子。该检测器对应于行人的身体轮廓（比如头，肩，脚）：(a) 所有训练数据得到的平均梯度图像；(b) 对于每个块 SVM 分类器的最大正权重（大的和重叠的块，位置在它们的中心像素）；(c) SVM 分类器的最大负权重；(d) 测试图像窗口样例；(e) 窗口 (d) 的方形 HOG 块描述子；(f)、(g) 窗口 (d) 的方形 HOG 块描述子，分别通过 SVM 的正权重 (f) 和负权重 (g) 进行加权。线性 SVM 分类器正确地将窗口 (d) 分类为行人。本图的彩色版见彩图 24



彩图 24 同图 10.32