

问题

在深度学习目标检测算法问世之前，HOG+SVM可以说是最经典的传统目标检测算法了，既然是经典，那自然需要了解了解对吧。所以今天我们就来看一看，HOG是如何进行特征提取的。

HOG简介

HOG（Histogram of Oriented Gradient，即方向梯度直方图）是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子，由Navneet Dalal和Bill Triggs等人在2005年CVPR发表的论文上提出。HOG通过计算和统计图像局部区域的梯度方向直方图来构建特征，结合SVM等分类器常用于图像识别等领域，尤其在行人检测中具有非常不错的表现。

HOG特征描述原理

• 主要思想：

在一幅图像中，局部目标的表象和形状（appearance and shape）能够被梯度或边缘的方向密度分布很好地描述。（本质：梯度的统计信息，而梯度主要存在于边缘的地方）。

• 具体操作：

先将图像划分为若干个cell，然后计算各个cell中每个像素点的梯度和边缘方向，然后分别统计每个cell的方向梯度直方图，构成了每个cell的特征描述子；

为了对光照和阴影有更好的不变性，需要对直方图进行对比度归一化，具体可以通过将局部直方图在图像更大的范围内（称为block）进行对比度归一化。组合每个block中归一化后的所有cell的descriptor，就构成了该block的descriptor。

最后将所有block的descriptor组合起来就是整幅图像的HOG特征描述子。

HOG特征检测步骤

1. 计算图像每个像素点的梯度大小和方向
2. 将图像划分成若干个cell
3. 统计每个cell的方向梯度直方图
4. 将相邻的若干个cell组成更大的block
5. 对每个block的进行对比度归一化
6. 将归一化后的所有block特征描述子组合在一起，构成整幅图像的descriptor

然后就可以将整幅图像的HOG特征向量结合SVM等机器学习算法进行训练，实现目标检测任务了。

通过上面的介绍，想必对HOG的原理已经有了大致的理解了，但对于过程中每一步具体怎么操作的是不是还有点模糊呢？不着急，我们接着往下看。。。

HOG算法具体实现

注意，有些文章或博客中HOG算法实现步骤中可能给出第一步是先进行图像预处理，如Gamma校正，即对图像的颜色空间进行归一化，其目的是调节图像的对比度，降低图像局部阴影和光照的影响，同时可以抑制噪声。但后来作者指出这一步其实可以省略掉，因为后面步骤中的block归一化处理能达到同样的效果，而图像预处理的贡献微薄，所以第一步可以直接计算梯度和方向。

1.梯度计算

计算图像上每个像素点的梯度大小和方向：

$$\text{水平方向梯度: } G_x(x, y) = I(x+1, y) - I(x-1, y) \quad (1)$$

$$\text{垂直方向梯度: } G_y(x, y) = I(x, y+1) - I(x, y-1)$$

每个像素点在 (x, y) 处的梯度幅值和方向分别为：

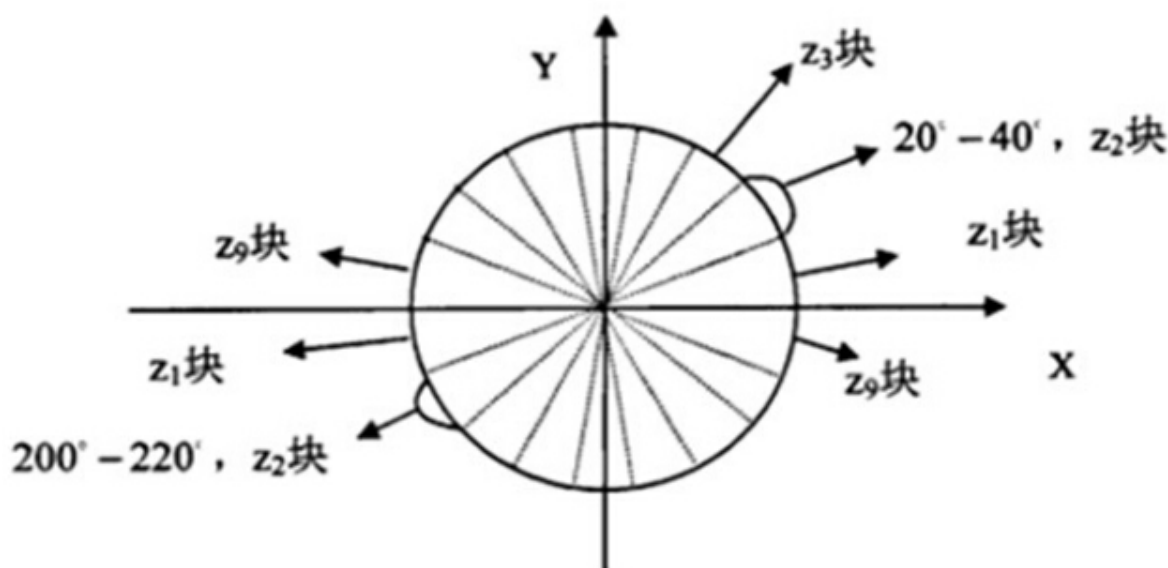
$$\text{幅值: } G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}$$

$$\text{方向: } \alpha = \arctan \frac{G_y(x, y)}{G_x(x, y)}$$

2. 计算梯度方向直方图

首先将图像划分为若干个cell，大小可以自由设定，例如将整幅图像划分为 8×8 个cell，每个cell都包含若干个像素。

然后分别统计每个cell的方向梯度直方图，大量的实验表明，统计直方图时采用9个bin来统计方向信息效果较好。具体来说就是，将cell的梯度方向 $0 \sim 180^\circ$ （或 $0 \sim 360^\circ$ ，即考虑了正负）分成9个方向块，如下图所示：



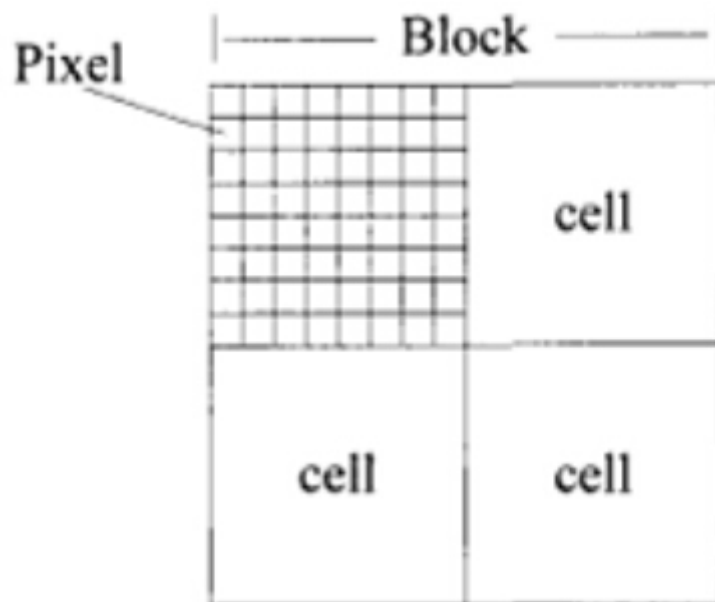
在直方图中，每 20° 组成一个bin，当梯度方向落在某个bin范围中，该bin上的计数就增加（可以理解为按照梯度方向来进行投影的过程）。那相应的bin计数应该增加多少呢？Dalal和Triggs等人认为，将梯度幅值大小作为投影权值效果通常较好。也就是说，假设某个像素点的梯度大小为3，其梯度方向刚好落在 $20 \sim 40^\circ$ 之间，那么该bin上的计数就增加3（也就是将梯度大小作为权值用来计数）。通过这样的一一映射，对于每个cell，都可以映射成9个bin的方向梯度直方图，即每个cell都是9维的descriptor。

3. block归一化

由于局部光照的变化，以及前景背景对比度的变化，会使得梯度强度的变化范围非常大，因此需要对梯度做局部对比度归一化。归一化能够进一步对光照、阴影、边缘进行压缩，使得特征向量对光照、阴影和边缘变化具有鲁棒性。

具体如何操作？

作者的做法是：将多个cell组成更大的block，然后对每个block进行对比度归一化。图像最终的HOG特征描述子就是所有block内的cell的直方图构成的向量。事实上，块之间是有重叠的，也就是说，每个细胞单元的直方图都会被多次用于最终的描述子的计算。块之间的重叠看起来有冗余，但可以显著的提升性能。



如上图所示，每个block由2×2个cell组成，每个cell包含8×8个像素，每个cell得到9个方向的bin，因此每个block的特征向量维度为2×2×9=36维。

作者采用的归一化方法有4种：（ $\|v\|_k$ 表示 v 的 k 阶范数， e 是一个很小的常数）

$$L2 - norm : f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \quad (2)$$

$L2 - hys$: 先计算 $L2$ 范数，然后限制 v 的最大值为 0.2，再进行归一化

$$L1 - norm : f = \frac{v}{\|v\|_1 + e^2}$$

$$L1 - sqrt : f = \sqrt{\frac{v}{\|v\|_1 + e^2}}$$

4.提取HOG特征

最后一步就是将归一化后所有的block进行HOG特征的提取，并将它们结合成最终的特征向量送入分类器。

那一张图像的HOG特征维度是多少呢？

例如，一张大小为64×128的图像，每个block由2×2个cell组成，每个cell包含8×8个像素，每个cell计算9个bin的方向梯度直方图，以8个像素为步长，则水平方向有15个扫描窗口，垂直方向有7个扫描窗口，所以整幅图像的HOG特征维度为15×7×2×2×9=3780维。

参考资料

目标检测的图像特征提取之（一）HOG特征 <https://blog.csdn.net/zouxy09/article/details/7929348>

第十八节、基于传统图像处理的目标检测与识别(HOG+SVM附代码) <https://www.cnblogs.com/zyly/p/9651261.html>