**《场景图像分类技术综述》阅读摘录**

1.场景图像分类的两个步骤：特征提取、分类

2.特征提取的概念：

将原始特征转换为一组具有明显物理意义（Gabor、几何特征[角点、不变量]、纹理[LBP HOG]）或者统计意义或核的特征。

3.场景分类演变的3个阶段（仅仅是演变趋势，无绝对优劣）：

（1）基于底层特征的方法：GIST、SIFT、CENTRIST、BOW、SPM、LLC、IFV

（2）基于高层特征的方法

①基于场景模型：ROI、MM-Scene、DPM、RBow

②基于场景语义：Object Bank、Patches、LPR、BOP

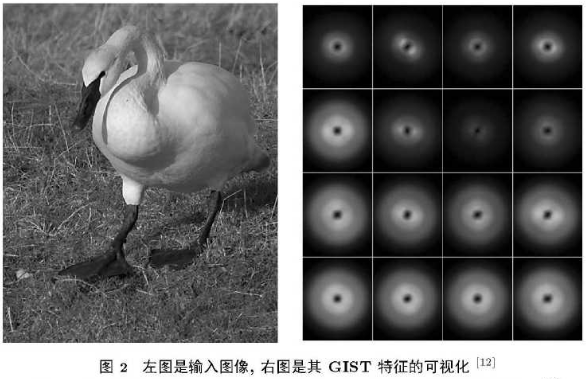
（3）基于特征学习的方法：CNN、SPP-CNN、MOP-CNN

4.底层特征

Meanning：以像素点为基础提取的图像原始属性，着重刻画图像细微的纹理信息

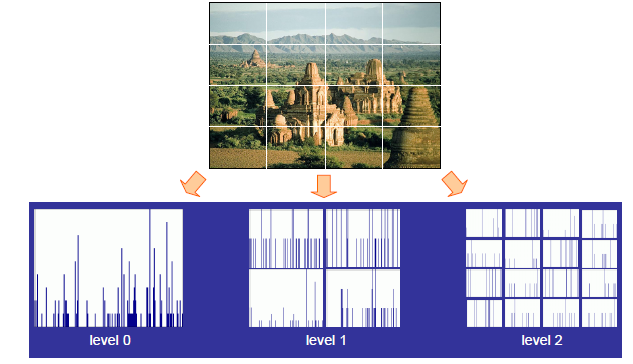
1. GIST特征（2001年，Oliva和Torralba）

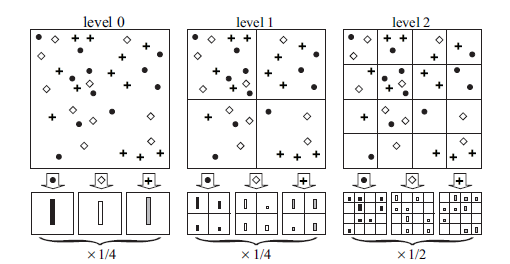
利用光谱信息反映图像的形状结构信息，反映图像的自然程度和开放程度等全局特征



1. SIFT特征（1999年，Lowe）

描述图像局部信息的强有力特征，图像划分，特征统计直方图，合理加权之后拼接成特征向量（SPM模型**Spatial Matching Scheme，2006年，Lazebnik**）

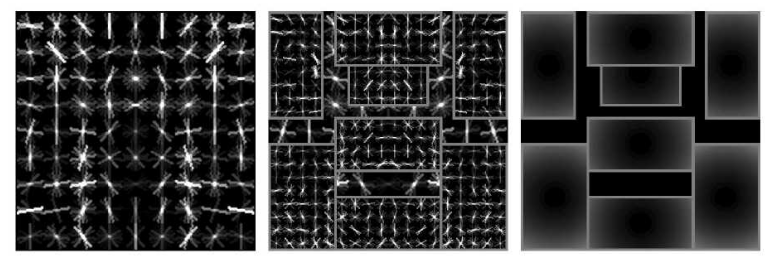




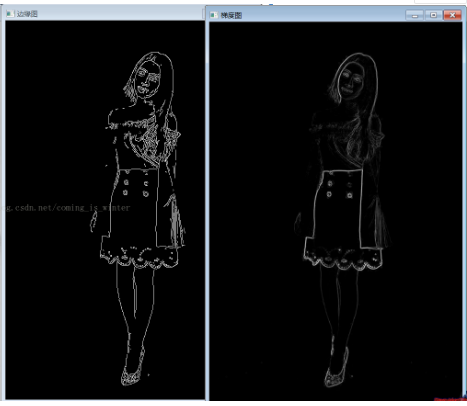
上面的黑圆点、方块、十字星代表一副图像上某个pitch属于kmeans后词典中的某个词；

1. 将图像划分为固定大小的块，如从左到右：1\*1，2\*2，4\*4， 然后统计每个方块中词中的不同word的个数；
2. 从从左到右，统计不同level中各个块内的直方图；
3. 最后个将每个level中获得的直方图都串联起来，并且给每个level赋给相应的权重，从左到右权重依次增大；
4. 将spm放入svm中进行训练和预测；
5. HOG （Histogram of Oriented Gradient）特征（2005年，Dalal和Triggs）

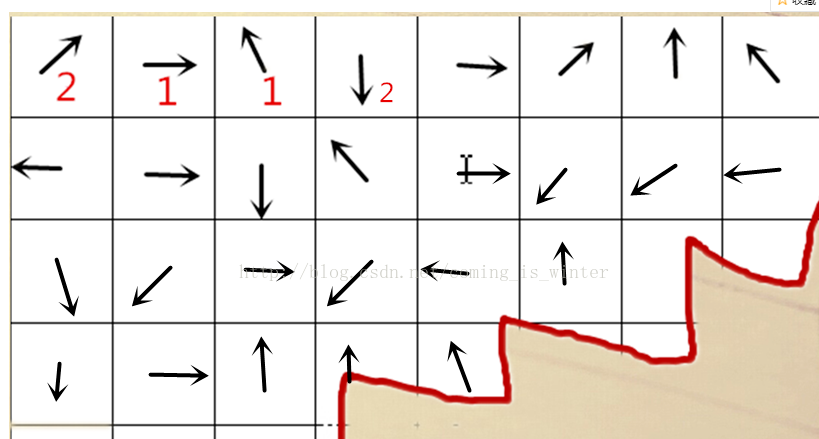
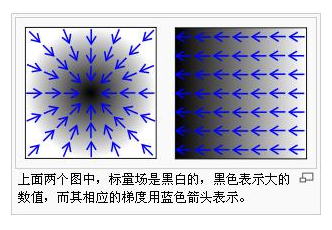
是一种能够捕获形状属性的图像特征，在一副图像中，局部目标的表象和形状能够被梯度或边缘的方向密度分布很好地描述。其本质为：梯度的统计信息，而梯度主要存在于边缘的地方。主要用于行人检测

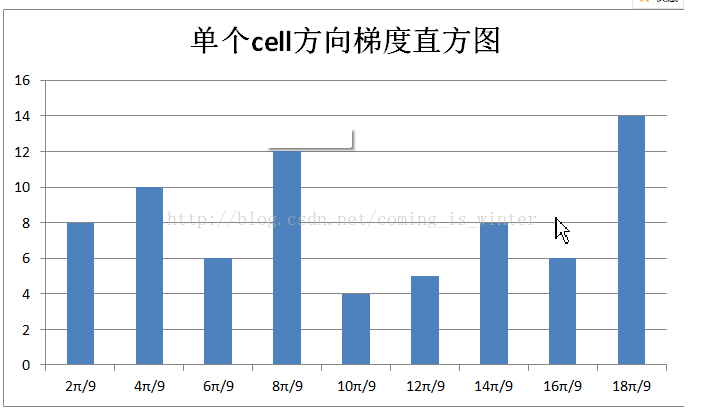






图像梯度的概念解释：标准色彩由亮变暗的方向趋势





之后再利用直方图，统计特征，进行分类比较

1. CENTRIST特征（2010年，Wu等）

描述图像局部形状结构的图像特征，像素点Census变换，得到整数值，然后统计直方图

Censua变换：将该点的灰度值与相邻8个像素点比较，如果中心点的灰度值大，则将对应的二进制位置为1，否则为0，形成一个八位二进制数，转成十进制，即Censua值。



（5）几种特征之间的比较

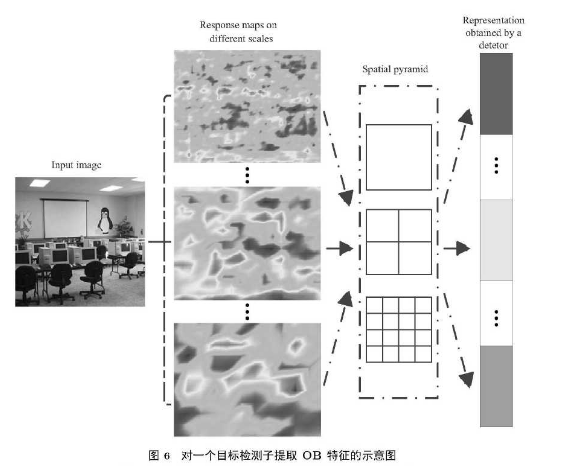
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 特征 分析 | 设计思想 | 应用场合 | 优点 | 缺点 |
| GIST | 全局特征，旨在通过局部光谱信息反映整体布局 | 捕获图像自然程度，适合应用在自然场景分类 | 特征计算简单易用，擅长整体刻画 | 在背景复杂或者目标密集场景表现差 |
| SIFT | 局部特征，主要通过统计图像关键点领域内的梯度方向来反映局部结构 | 能考虑到目标位置信息，对大部分自然场景和简单的室内场景都有不错效果 | 对平移缩放，尺寸变化，亮度变换，视角变换有一定的稳定性 | 不能描述复杂场景的整体布局 |
| HOG | 局部特征，侧重于统计图像像素块的边缘地图方向信息刻画目标的形状结构 | 能够捕捉几何结构和整体布局，对于稳定的全局结构有突出分类表现 | 能对边缘形状，目标轮廓进行描述 | 在场景形状结构不稳定的情况下会失败 |
| CENTRIST | 局部特征，通过统计图像块内Census变换值来反映区域的整体结构信息 | 通过局部结构信息刻画整体结构，适用于有清晰布局且不太密集的场景分类 | 通过局部信息得到几何位置信息的场景全局表示 | 对于缩放，旋转不具有不变形，且忽略了颜色信息，在复杂多变场景表现不好 |

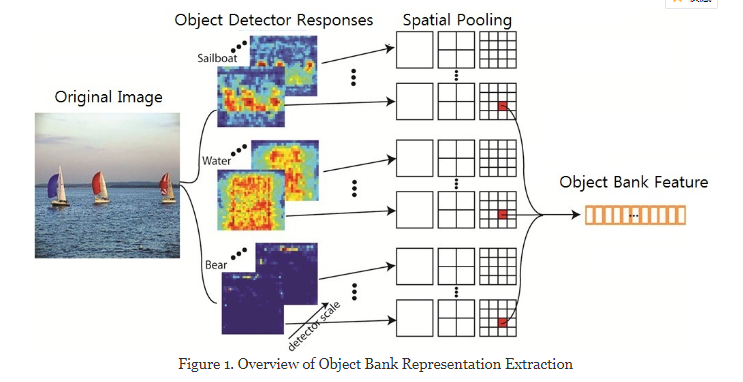
5.高层特征

Meanning：以底层为基础通过建模提取出来的高层属性，蕴含丰富图像语义信息

1. Object bank（OB）特征向量（2010，Li等）

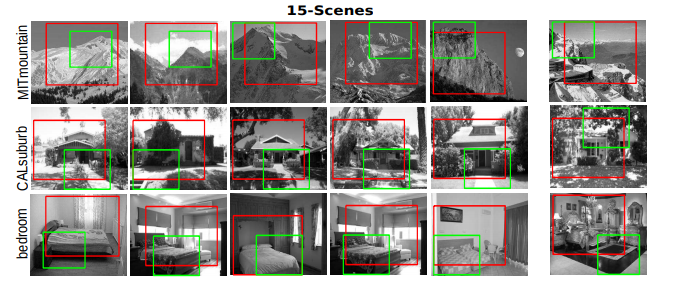
Object Bank在12个detection scales上使用了200个object detector，有3个spatial pyramid level（L=0，1，2）。这样就产生了No.Objects x No.scales x (1^2+2^2+4^2)个grid。对于每个grid，计算它的maximum response。对于每一个grid就产生了长度为No.objects的特征向量。将这些特征向量串联起来就得到了Object Bank对于一个图像的表示。

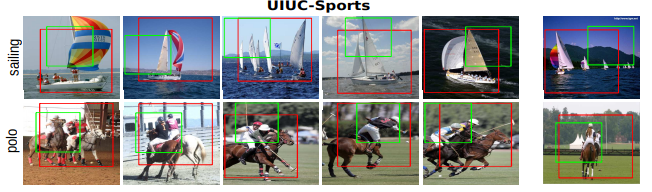


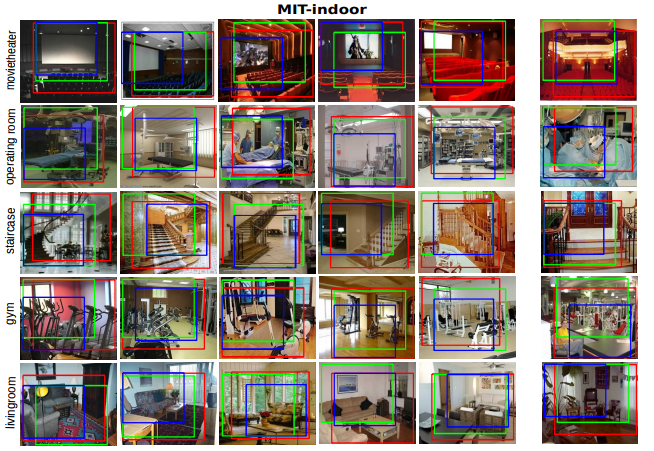


1. Latent pyramidal regions（LPR）特征向量（2012年，Sadeghi）

先用Latent support vector machine（LSVM）训练出若干区域检测子，构造LPR向量。对于一幅输入图像I，假设其LPR表示向量为v，则向量的每个元素都是一个特定监测子的模型参数和窗口特征。





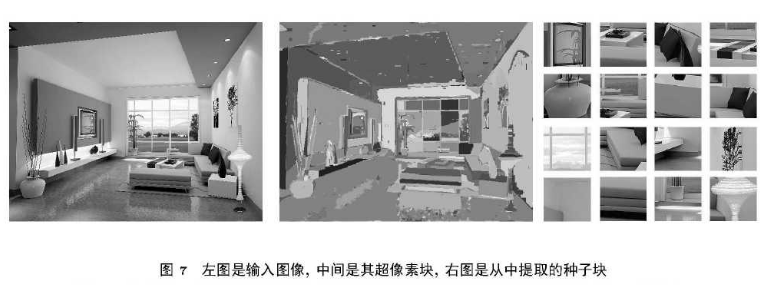


Latent Pyramidal Regions for Recognizing Scenes（Fereshteh Sadeghi and Marshall F. Tappen）

1. Bag of parts（BOP）特征向量（2013年，Juneja）

通过分割的方法，将图像划分成超像素块，进行筛选后，划分成HOG胞元，生成检测子，检测多尺度响应图，再按照空间金子塔模型，求最大响应值，形成特征向量。

（提取了检测子后，步骤和OB很类似）



1. 高层特征分析与比较

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 特征 分析 | 设计思想 | 应用场合 | 优点 | 缺点 |
| OB | 通过多目标检测子来识别场景包含目标的高层特征，目标确定场景类别 | 根据目标种类和位置来识别场景，适合一些有明显标志性目标的场景，背景不能太复杂 | 理解深层语义的基础上分类，思路与人类类似，大部分自然场景和简单室内场景比较有效 | 计算复杂度和特征维数字比较高，算法性能严重受限与所选取的目标检测子 |
| LPR | 借助区域检测子来描述场景特定区域的语义特征，特定形状和结构的区域确定场景类别 | 通过场景的显著性区域对场景进行辨别，理论上适合所有场景，尤其适合复杂拥挤的室内场景 | 对场景中具有特定形状的结构的区域能够很好的描述，适合复杂场景 | 缺乏对场景的深层次理解 |
| BOP | 通过超像素分割的方式来识别并刻画场景显著性区域的高层特征，特定形状和结构的区域确定场景类别 | 通过场景的显著性区域对场景进行辨别，理论上适合所有场景，尤其适合复杂拥挤的室内场景 | 擅长刻画场景中的目标、边界、棱角，适合复杂场景 | 缺乏对场景的深层次理解 |

6.特征学习

在机器学习任务中，将图像像素作为系统输入，经过一系列映射和转换，学习到潜在特征的过程（有监督，无监督），CNN便是一种，基于监督的深度学习。

7.分类

（1）基于常规分类器分类

SVM、KNN、Bayes、BP神经网络，利用PCA降维

（2）基于场景模型分类

DPM、ROI

（3）基于深度学习分类

CNN