Real-World Material Recognition for Scene Understanding

一 摘要：

本文我们解决日常用品图片的材质识别问题。尽管材质识别不是一个新问题，opensurfaces 数据库的引入使材质识别被更多的人研究。这一数据库提供了大量的各种各样的显示生活中的材料，每一种类的外观也大不相同。我们提出一种在图像中逐像素的的材料分类的可区分的学习模型。材料外观的差异大使得材料的分类更加困难，我们的方法仅仅取得了 34.5%的正确率。但我们显示了即便这种很弱的材料信号对场景分析也是很有用的。我们把我们分类出的材料信息作为 RGB-D 场景语义分析的一个新特征，这使我们的场景分析提高了 0.7%。

二 引言：

物体的材料信息有可能成为我们理解场景的重要信息。从一个简单物体中，我们可以判断它的高度，纹理，并通过鉴定它的材质组成判断它的功能。如图一中所示，我们都会把他们标记成瓶子，然而，通过判断它们的材质组成，我们可以将它们按重量分类，判断哪个会摔碎，或者哪个能装很热的水。这些属性的认识使材质识别成为场景分析一个很有意义的方向。尽管这样，材质识别在场景理解领域还是很少被研究。

图一：三个外观相同，物理性质不同的瓶子。

这项工作的研究直到 Bell 提出一个现实中的大规模材料库才被激活。这一数据库包括了 10000 张现实中的图像，每一张被由表面的反射，颜色，粗糙度和上下文信息分割和标记了。尽管纹理识别中还有一些其他数据库，自然环境下的材质识别仅仅是最近才被研究，没有哪个数据库能像 opensurfaces 那样更好地描述现实世界中的场景。

我们希望场景中材料组成信息的理解有助于场景理解，这需要对图像中的每个像素进行标记。因此，我们的方法寻求图像中逐像素的材料标记，然而Sharan 在[Material perception: What can you see in a brief glance?] 及其后续工作中仅仅识别图像中的主材料。

我们的方法建立在[Robust higher order potentials for enforcing label consistency]和 [Graph cut based inference with co-occurrence statistics] 提出的模型上。这种方法是图像上的条件随机场（CRF）来确保分割是光滑的，并且是连续的标记，但本文中我们只会考虑它的一元输出。使用随机森林方法进行材料的分类，我们将提取出四种特征：SIFT,COLOR SIFT,LBP 和Textons。

三 相关工作：

[Reflectance and texture of real-world surfaces]中使用CUReT数据库，并将纹理考虑进来。他们对表面纹理的视觉属性通过早期的纹理识别方法来描述。然而，这一数据库是在特定条件下生成的，只包括二维的纹理块，因此很容易实现。[A statistical approach to material classification using image patch exemplars] 中在CUReT数据库上取得了95%的正确率，但在现实中的Flickr 数据库只取得了23%的准确率。

[Exploring features in a bayesian framework for material recognition] 中使用很多材料分类的新特征，包括沿着图像边缘的HOG特征。然后经过k-means算法将这些特征聚类成视觉词典，然后使用贝叶斯模型将他们连接起来，这一方法在Flickr图像库上取得了45%的正确率，但每张图片只需要一种材料，因此很适用于场景理解。

[Toward robust material recognition for everyday objects] 在Flickr数据库上取得了54%^的正确率，它使用[Kernel descriptors for visual recognition] 中的kernel 描述子。他们也把特征量化成视觉词汇，但使用大间隔最近邻居算法分类。

四 方法：

我们的方法是建立在ALE模型上，它使用CRF（条件随机场）。本文我们主要集中于对图像的逐像素的材质识别上.

特征:

我们的算法以四个向量作为输入来对图像进行逐像素的材质识别，它们是SIFT, Color SIFT，LBP和Textons，它们的描述如下。

1） SIFT 和Color SIFT

SIFT 是由David Lowe十年前提出来的，现在是计算机视觉领域广泛使用的特征描述子。只有不同尺度在特征点才会被提取出来，然后计算这些点周围像素梯度。通过把梯度方向旋转为主方向来使它保持旋转不变性，最后生成一个128维的特征描述子。

Color SIFT是SIFT特征的一个变体，它包含颜色不变性来使它对图像中的颜色变化更鲁棒。SIFT是在灰度图像中提取特征点，Color SIFT是在颜色空间。我们用4个尺度，8个方向来计算SIFT和Color SIFT。

2） Textons

Textons 是最早发明用来纹理识别的特征描述子之一。Leung 通过将图像与n个滤波器组卷积，最后每个点生成一个n维的向量作为表示。这些向量然后通过k-menas聚类成视觉词汇，每个像素被映射到最近的邻居来生成一个基元图，本文中我们使用150个聚类中心。

3） LBP

LBP是由Ojala提出，它提供另一种在像素周围表达图像结构的方法。对于一个特征点周围的区域的每个像素，我们使用一个8位的向量来记录哪个像素的8邻居变化小。每个区域的这些向量然后被统计成一个直方图，这就是LBP特征。同样的，对所有图像的LBP特征，我们使用k-means算法进行聚类。

训练：

本文我们使用随机森林算法，这种算法被证明在实时数据，人类pose分类和边探测上取得了很好的效果。这一算法学习了一系列的决策树，每一棵决策树是特征集的随机子集。在树的每个节点，我们从树的随机抽取的特征集中选择最好划分数据的特征。每棵树的叶子节点是类的分布。在分类阶段，像素的特征向量是由每棵树在像素的类分布上投票进行分类的，这些分布平均寻找最终的分类概率。一棵但意思的树通常受到差异大的影响，这种方法可以通过把树添加到森林中来降低这种影响。我们的实验中采用不同大小和深度的森林。

尽管数据库中有53种材料分类，但很多图像没被标记好，还有些不常在场景分析中出现。

我们从中抽取16个分类，每个分类随机抽取1000张图片。我们去掉一些太小的样本（长度或宽度小于30），这样剩下了12568张图像。一半用于测试一半用于训练。

五 实验结果

我们训练的随机森林有10,20,30,40,50棵树，5,10或15个节点深。训练的随机森林的树的个数大约是线性的，深度是指数型的。训练50棵树，15个节点深的森林在24核的电脑上大概花费了一周的时间。训练20个节点的单棵树用了大概24个小时。图2 是随机森林的森林个数和深度对分类正确率的影响。

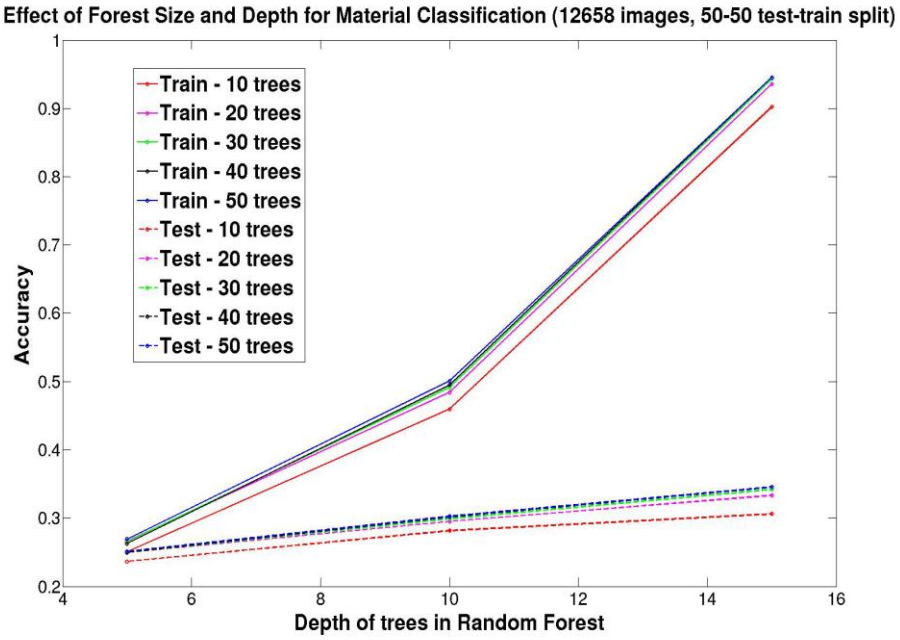

图2：不同大小和深度的测试准确率

图3 是随机森林为50棵树，15个节点深的分类正确率的混淆矩阵。这一算法对测试集的正确率大概是34.5%，在训练集上是94.6%，两者的巨大差异表明算法过拟合严重。

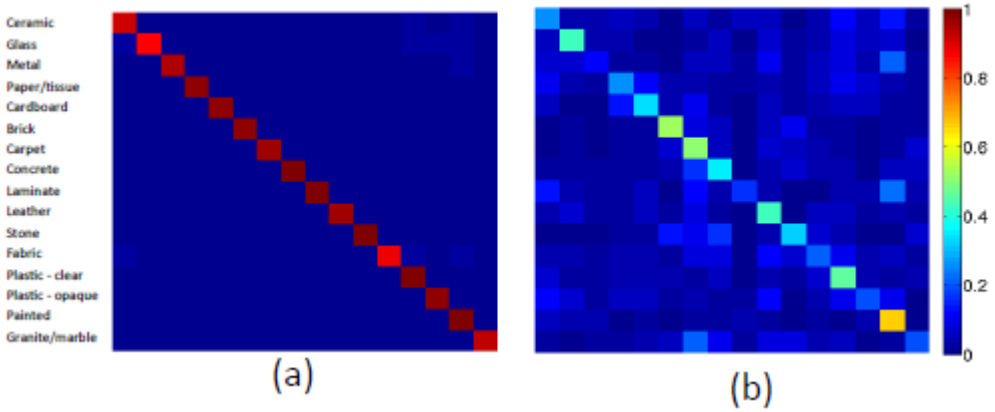

图3：50棵树，15个节点深的随机森林在训练集（a），测试集（b）上分类的混淆矩阵