

Texture Classification

引言：

我们研究从材料图像的外观上识别材料的问题，不需要添加任何限制，任何对材料的先验知识和光照条件。这个问题可以描述成：给定一张单一的材料纹理图像，把它分类到我们知道的类别当中。

从一张普通的图像中识别它的纹理是很有必要的。有别于其他形式的分类，分类物体有给定的形状来获取，大多数纹理有各种各样的多样性使得很难对他们进行建模。还有，纹理图像经常受光照，拍摄角度等因素干扰，解决这一问题是什么分类算法需考虑的问题之一。

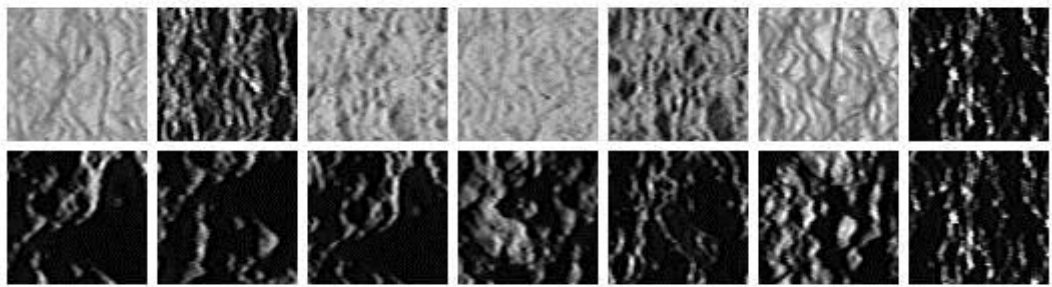


图 1：同一类别的纹理由于拍摄条件造成的多样性。上行：特定视角不同光照，下行：特定光照不同视角。

还有一个问题是两种不同材料的图像有时候也可能很相似。这些问题都加大了纹理分类的难度

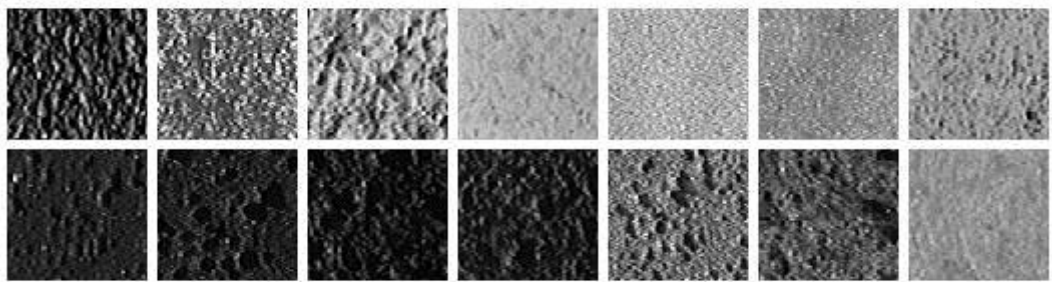


图 2：类间的差异性很小的表现。上行：第一张和第四张是同一类别的，其他的术语不同类别但看起来相似，下行：三种不同类别的图像外观看起来很相似。

弱分类算法，通过把图像表示成滤波响应或图像块的统计分布将成为我们研究的基础。

实验准备：

CURet 图像库包含了 61 种材料，基本上涵盖了我们日常所见的，它包含了粗糙的，有反射光的，人造的等各种纹理。下图是图像库中的一些纹理图片。

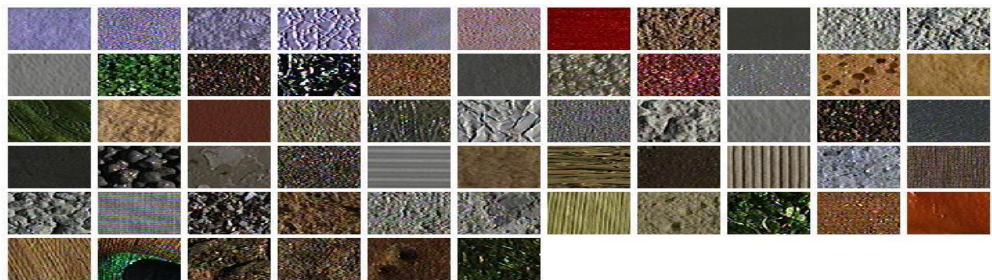


图 3：从 CURet 图像库中每个类别抽取的一张图像。

图像库中每个类别的图像都是在 205 种不同视角，光照下拍摄的。图像的相互反射，阴影和材料表面的多样性使得在 CURET 图像库上的材质识别比 Brodatz 图像库难得多。

但是，CURET 图像库还是有一些局限性。主要是图像拍摄的方式和纹理的选择。对于前者，大多数材料没有太大的尺度改变和平面的旋转。对于后者，最大的缺陷是有些材料，相同的纹理有多种不同的表现形式，因此类间的差异性很难被区分。然而，CURET 始终是纹理分类里最大，最全面的一个数据库，因此我们选择这样图像库用来研究。

对于图像库中的每种纹理，图像库都有 118 张方位视角小于 60 的图像。我们从每个类中选取了 92 张充分大的纹理图片。我们选取中心区域 200*200 大的区域，然后将其转化成灰度图像进行归一化。因此我们的实验中没有颜色信息。下图是木材类的 92 张被选择的图像。

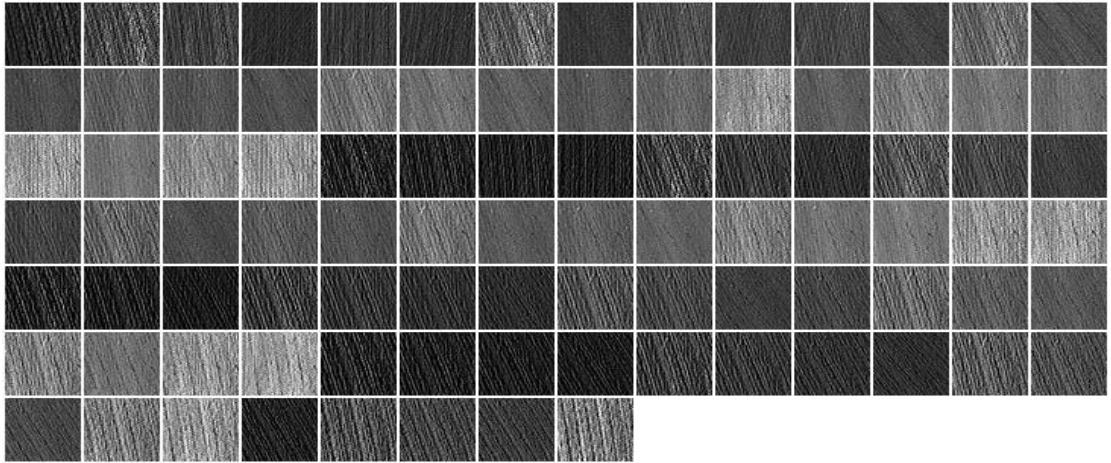


图 4： 木材类的 92 张剪切的图像

因此，我们的图像库中有 $61 \times 92 = 5612$ 张图像，它们被平均分成了两类，一般用于训练一般用于测试。

有过滤的分类：

滤波：

我们的算法，跟通常的弱分类器一样。分为训练和分类两个阶段。在训练阶段，训练图片经过滤波后生成滤波响应（filter responses）。这些滤波响应通过 k-means 聚类后生成基元（textons），所有的基元构成一个字典。

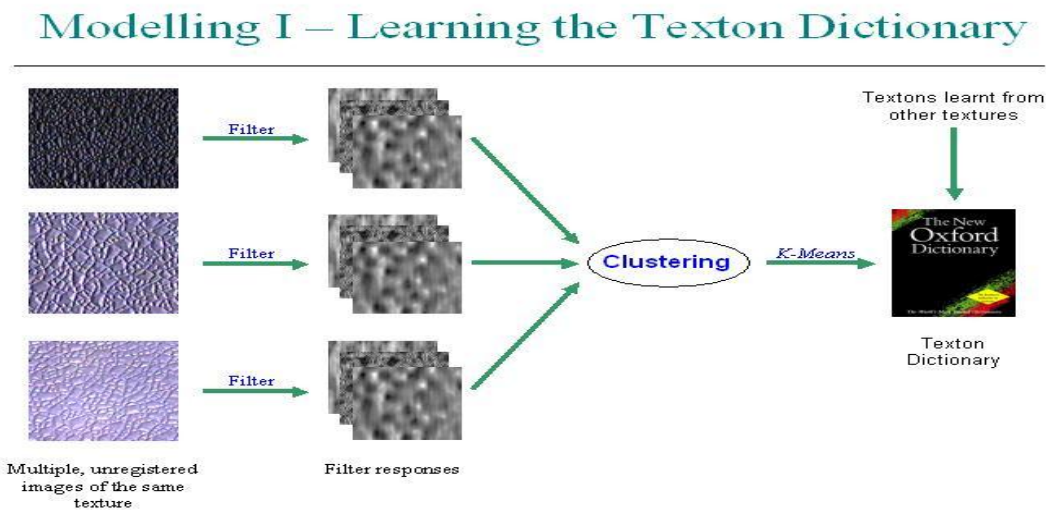


图 5： 生成基元字典（textons dictionary）

我们从训练集的每个类别中随机抽取 13 张图片，这些图片是在不同光照和视角下的。这些图片的滤波响应随后被聚类成 10 个中心，这些中心就是基元（textons）。每个类别的基元然后聚合在一起形成基元字典，比如有 20 个类别基元字典就会有 200 个基元。

下一步就要为每个类别建模了。对于每一张训练图片，首先通过滤波器卷积，然后对每个滤波响应选择离它最近的基元（texton）表示，这样就生成每张图片对应的模型。基元的直方图，每个基元出现的频率就是训练图片对应的模型。每张图片就产生一个模型，因此每个类别就有很多个模型来表示。

Modelling II – Multiple Models Per Texture

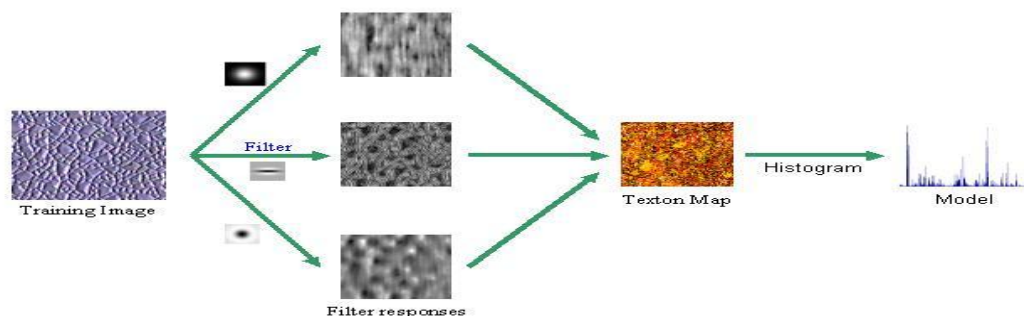


图 6： 为训练图片建模

在分类阶段，我们采用同样的方法来表示新图片，生成的直方图来对训练图片的直方图进行比较。我们使用卡方距离来衡量距离，使用最近邻算法进行分类。

Nearest Neighbour Classification

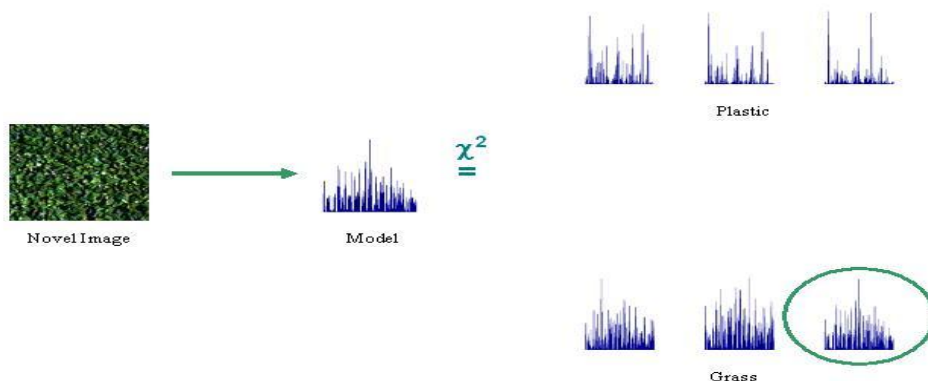


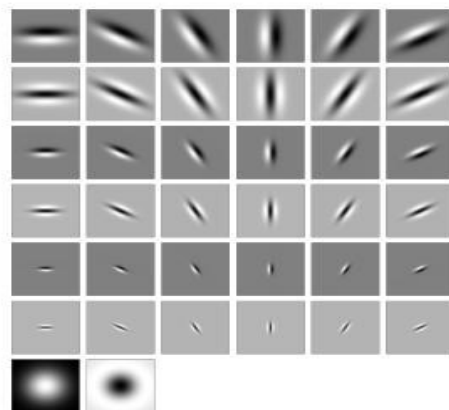
图 7： 图像分类

滤波器：

我们的滤波器将和这些作比较：Leung 和 Malik 的不是旋转不可变；Schmid 的；还有源于 RFS 的 MR4, MRS4 和 MR8 滤波器。这些滤波器将会由他们的滤波响应上的分类效果进行比较。

MR8 滤波器组有 38 个滤波器但只有 8 个滤波响应。但他们的输出只会记录最大的滤波响应方向。这使得滤波器确保了旋转不变性。MR8 滤波器组如图 8 所示，它包含一个高斯，高斯拉普拉斯滤波器，这两个都有旋转对称性，还有 3 个尺度下的边缘滤波器和 bar 滤波器，后面两种在每个尺度有 6 个方向。只考虑最大方向的滤波响应将滤波响应从 38 减少到 8。

Maximum Response Filters (ECCV 2002)



Maximum Response (MR8) filters – 38 filters, 8 filter responses

图 8: MR8 滤波器, 低通, 旋转不变的, 同时也会采集方向特征。

实验结果:

我们算法的运行结果是由 CURET 图像库的不同材料的分类来衡量的。我们做了 3 个实验来研究材料种类的个数对实验结果的影响。我们从 20 个类别中学习出一个 200 大小的基元词典。实验结果如下表所示。

Filters	Dimension	Invariance	Number of texture classes		
			20	40	61
S	13	Rot.	96.30%	95.27%	94.62%
LMS	48	None	96.08%	93.75%	93.44%
LML	48	None	98.04%	96.47%	96.08%
RFS	38	None	98.37%	96.36%	96.08%
MR8	8	Rot.	97.83%	96.41%	96.40%
MR4	4	Rot.	94.13%	92.07%	90.73%
MRS4	4	Scale, Rot.	96.41%	94.08%	93.26%

表 1: 不同材料类别数对分类效果的影响。

下图显示了一些其他因素对 CURET 图像库分类的影响, 左图显示了基元词典个数对分类准确率的影响, 当基元词典为 2440 的时候取得了一个最大的准确率 97.43%。右图画出了当每个随机抽取的类别 46 个基元的分类正确率的分布情况, 平均分类正确率是 97.28%, 标准差大概是 0.316%, 最高为 98.4%, 最低为 95.72%。

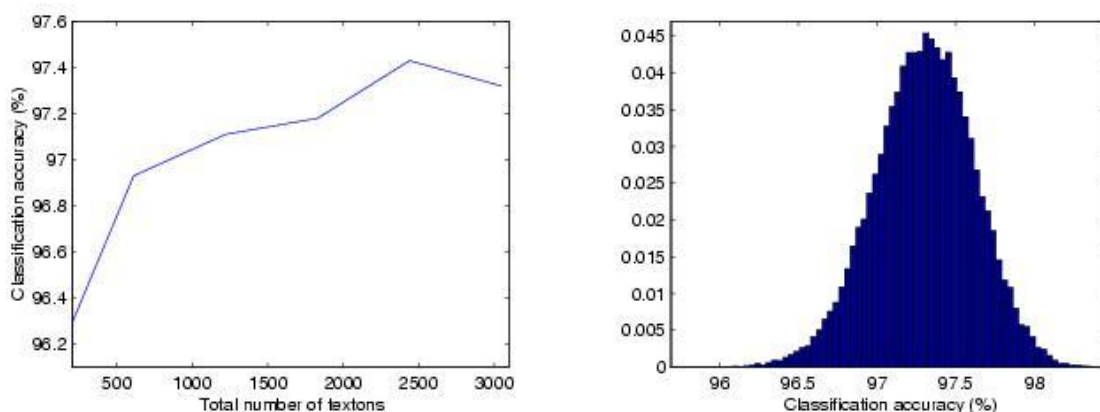


图 9: 左: MR8 分类效果与基元词典个数的关系, 右: 每个类别 46 个基元的分类正确率的分布情况

无滤波分类:

我们来研究大量的滤波器在纹理分类中扮演什么角色。我们的算法主要集中在使用滤波器的纹理分类, 并且表明是否用像素点和它固定周围的形成的块代替滤波器的滤波响应能取得更好的分类效果。

基于图像块的分类:

基于图像块的分类算法是建立在有滤波器的分类上的, 唯一的不同是在滤波阶段, 我们不使用 MR8 滤波器来逐像素生成滤波响应, 我们使用像素点周围 $N \times N$ 的区域, 在这一区域内逐行排序来生成向量。其他的所有操作都没改变。这种分类器也被叫做连接分类器。

为研究图像的马尔可基夫性质, 我们也考虑邻居分类器。它和连接分类器很相似, 只是中心像素不再包括在特征表达上了。因此, 图像块的向量是在 $N \times N - 1$ 的空间里通过逐行的排序生成的。这种分类器可以表示中心像素和它的邻居的分布紧密情况。

最后, 马尔可基夫随机场分类器确定中心像素在其邻居的 PDF (概率密度函数)。因此, 我们不仅不忽略中心像素, 还要把它的 PDF 通过量化成 bins 存储起来, 邻居则照样转化成 textons。因此联合概率密度是一个 $N_t \times N_b$ 的矩阵, N_t 是存储邻居 PDF 的基元个数, N_b 是存储中心像素的 PDF 的 bins 的个数。矩阵的每一行是中心像素在给定大小的邻居内的 PDF。实际上, 我们要建模的是纹理的马尔可基夫随机场分布, 除了那些已经被使用过的结合点。这样提高了分类的准确率。

Learning / Classification Without Filtering

Model changes from joint PDF of filter responses to joint PDF of raw pixel intensities computed over all $N \times N$ patches in image:

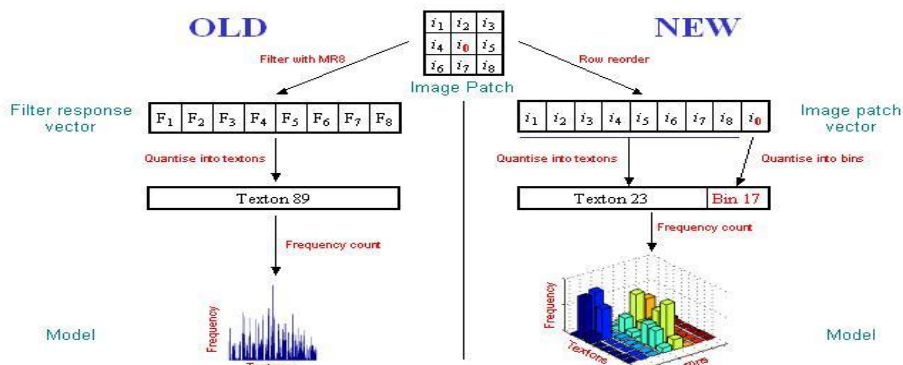


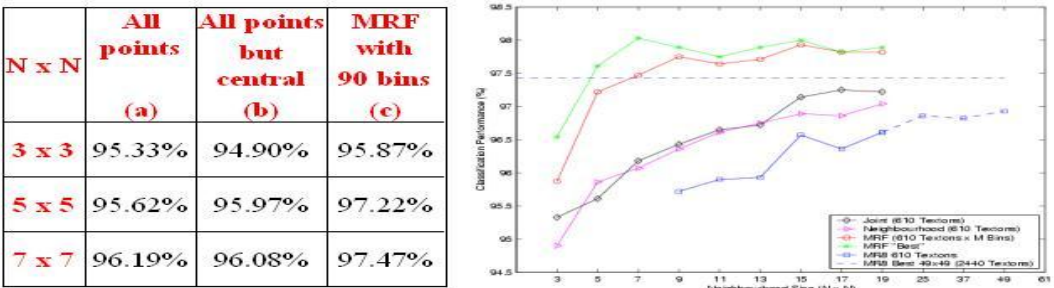
图 10: 马尔可基夫随机场表示

实验结果:

算法的分类结果随邻居范围的变化如图 11 所示。图像块小比多尺度, 多方向的滤波器效果更好。最

好的滤波效果也不如固定大小的图像块效果好。

Results

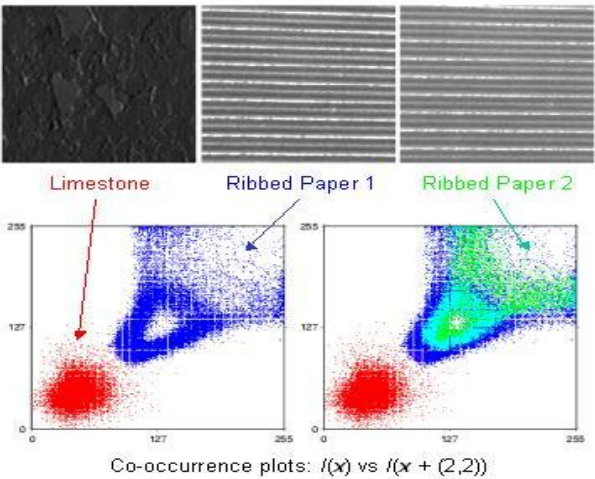


- Extremely surprising - 3×3 outperforms MR4, S and LM
- MR8 always worse for every comparable neighbourhood size
- Best – 98.03% (7×7 , 2440 Textons, 90 Bins)

图 11： 图像块大小与分类效果的变化关系

实验结果表明基于马尔可夫随机场的分类算法比基于滤波器的算法分类准确率更高。全局结构的纹理的大小远大于用于分类的局部邻居的大小。

Why Should Small Image Patches Work?



Distributions are clearly distinct $\Rightarrow 3 \times 3$ patches are sufficient for classification

图 12： 3×3 大小的块已经足够用于分类

两张纸的图像分布可以很容易地相互联系也可以很轻易地和石头图像区分开来。因此 3×3 大小的图像块包括足够的信息用于区分尽管他们的全局的纹理结构比局部快大得多。

然而，还是存在一些不能仅仅基于局部信息区分的类别。比如两种有相同的一阶基元统计但高阶同一不一样的纹理就很难区分。使用 7×7 大小的图像块能得到 98% 的正确率这一事实说明 CURET 图像块不属于这种类别。

尺度，旋转，合成：

CUReT 图像库缺少尺度变化可能是造成 MRF 分类器比 MR8 效果更好的原因。为证明这一假设，我们选择了 4 个有尺度缩放图像类别，并用以上方式分类。下图是分类效果：

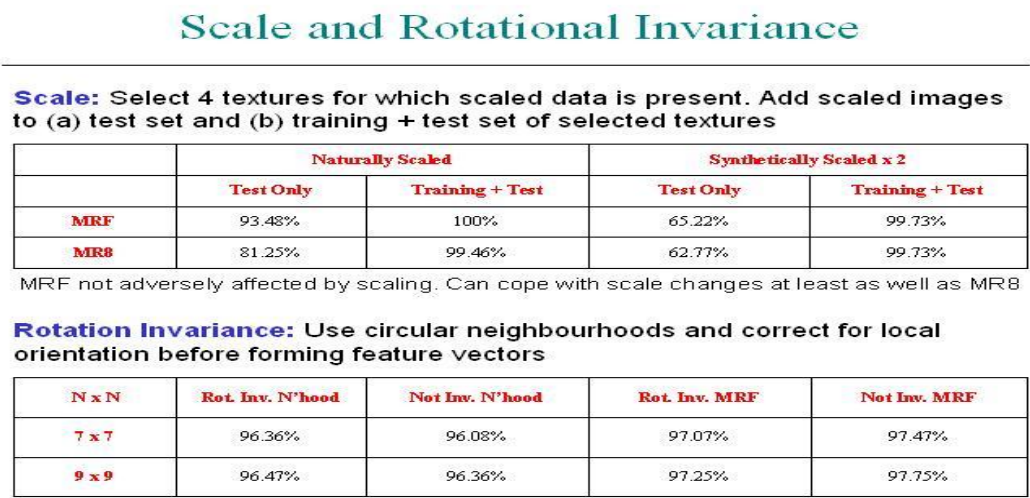


图 13：尺度变化与旋转的效果。

我们也可以通过获取每个块的方向并在生成特征向量之前进行修改来使我们的特征表达具有旋转不变性。下图是合成后的效果：

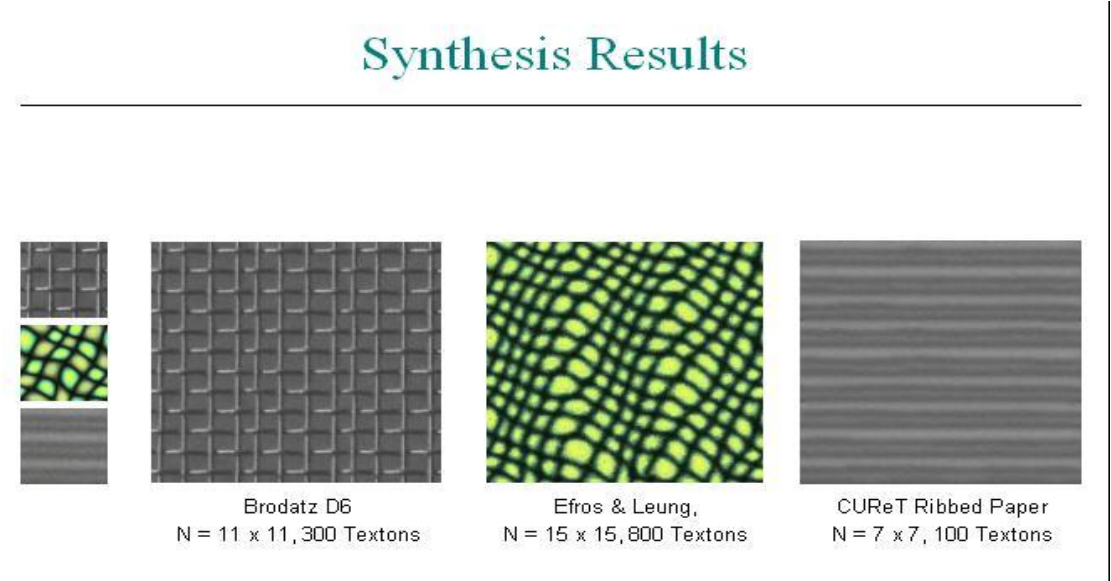


图 14：特征表达的合成效果。

我们图像块的联合概率分布的表达也同样适用于合成。合成算法与 Efros 和 Leung 的很相似，有两个参数。第一个参数是定义邻居的大小，第二个是基元词典的个数。中心像素的 PDF 我们使用 256 个 bins 来存储。

总结：

比起小的图像块，滤波器有很多不足之处：第一，图像块小，从训练集得到的信息也更多（一张图像中 3*3 块的个数比 50*50 块的个数多得多）。其次，图像块越大在图像分割中也是不利的，由于滤波两个区域的边界也是不清晰的。再次就是滤波使得很多局部信息丢失了。最后，从所有输入数据的学习到的分类器要优于从低维滤波表达学习到的。