

# 图像视频大作业二

2014011355 辛杭高

2017 年 5 月 11 日

本部分主要是对给定数据集进行检索工作，采取了不同的图象特征提取方法，也采用了不同的距离向量，对各组合下的结果进行了对比与分析。

## 1 Usage

本实验的全部代码由 python 编写，依赖库如下：

scikit-image

opencv(cv2)

numpy

PIL(pillow)

运行时必须保证 AllImages.txt 和 QueryImages.txt 与 main.py 在同一目录下，运行命令为

```
python main.py feature_choice distance_choice
```

如：

`python main.py 242 L2` 表示用 242 的颜色划分来获得直方图，在比较直方图的时候使用欧式距离。

运行得到的结果会保存在 result 文件中，如 `result/L2 242` 表示 242 颜色划分下使用欧式距离比较直方图得到的分类结果。

其中颜色的划分可以从 [242, 484] 两种中进行选择，距离的度量可以从 [L2, H1, Bh, ch] 四种中进行选择。如果遇到无法运行的问题，请联系 [xhg14@mails.tsinghua.edu.cn](mailto:xhg14@mails.tsinghua.edu.cn)。

## 2 直方图

直方图用于描述图象在 RGB 空间上的特征，其中 242 表示将红色和蓝色分为两部分，将绿色分为四部分。同理，484 表示将红色和蓝色分成四部分，将绿色分成八部分。由于时间复杂度和空间复杂度的限制，我们不可能将颜色空间分得太细，否则复杂度会直线上升。

显然，直方图只能描述图象的 RGB 特征，对于颜色在图象上的位置对直方图的生成完全没有影响，这也是直方图的一个局限。

## 3 直方图距离的度量

在图象检索任务中，我们将一张图的颜色直方图作为其特征。将两张图片的相似性归于两个直方图的相似性。为了度量直方图之间的距离，我们引入了 4 种距离度量策略。

### 3.1 Euclidean(L2)

$$d(x, y) := \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

这种策略将直方图的各 bin 的权重看成向量，比较两个向量之间的距离。

### 3.2 Histogram Intersection(HI)

$$d(x, y) = 1 - \frac{\sum_i \min(x_i, y_i)}{\sum_j y_j} \quad (2)$$

可以理解为两个直方图相交的面积，HI 的值越大，图象的相似度越高。

### 3.3 Chi-Square(ch)

$$d(x, y) = \sum_i \frac{x_i - y_i^2}{x_i + y_i} \quad (3)$$

颜色划分	Bh	ch	HI	L2
242	0.3975	0.3802	0.3444	0.3037
484	0.4889	0.4778	0.4568	0.3815

表 1: 传统方法下各参数对分类准确率的影响.

对应于 distance.py 中的 ch 函数。

3.4 Bhattacharyya(Bh)

$$d(x,y)=\sqrt{1-\sum \sqrt{x_iy_i}}$$

(4)

对应于 distance.py 中的 Bh 函数。

4 实验结果以及分析

4.1 距离度量分析

显然，由表 1 中的结果可知，在本测试集上，距离度量的准确率排序为 Bh > ch > HI > L2。



对于 L2 距离来说，它仅仅考虑到了绝对距离，而没有考虑到相对距离。比如上述中的 3 张图象，L2 会认为 A 与 C 的距离要近于 A 与 B 的距离，而其他算法给出的结果都相反。这主要是因为图象 C 中有大量的绿色所以与 A 图在绿色上的欧式距离差距比较小。但是 C 中出现了 A 中没有出现过的红色，所以 C 图的红色分量相对于 A 的红色分量距离会很大，所以针对考虑相对距离的度量来说，A 与 C 的差距才能被更好得度量出来。A，B，C 三张图在各个距离度量下的结果如表 2 所示

度量方法	A 与 B	A 与 C
L2	0.0998	0.0894
HI	0.9031	0.8936
ch	0.1122	0.1610
Bh	0.1546	0.1608

表 2: A,B,C 在各个度量下的结果.

Histogram Intersection 可以看成两个直方图相交的面积,一定程度上能反映两个直方图的相对误差。但也有一些比较特殊的情况,比如对于向量  $[0.4, 0.4, 0.4]$  和向量  $[0.6, 0.3, 0.3]$  在 HI 度量空间下,两向量与  $[0.3, 0.3, 0.3]$  的距离相同,但是在 L2 度量空间下,前者的距离要更小一些。显然,在这种情况下 L2 的判断更为合理一些。

Bhattacharyya 度量方法可以看成是 Histogram Intersection 的升级版,它利用乘法来考虑两个直方图的重合部分,如果直方图的某个 bin 的值很小,则必然会增大 Bhattacharyya 度量值。当然 Bhattacharyya 度量要求被度量向量  $x, y$  各项和为 1。

Chi-Square 是统计学上度量两者相似性的常用策略,在样本数目越多该相似性度量越准确,对应于本实验来说,当颜色空间 bin 的数量取得很多的时候,Chi-Square 统计会有很好的效果。

## 4.2 颜色空间划分

无论是 242 的划分方式,还是 484 的划分方式都选择将绿色划分地更细致一些,这主要是考虑到人眼对于绿色更为敏感,所以将绿色划分得更细一些有助于提升人眼对图象的辨别的能力。但是最终进行辨别任务的并不是人眼,而是距离度量函数,因此我对于将绿色划分得更细是否有用表示怀疑。为此,我在 4 种距离度量方法下测试 422, 242, 224 三种划分策略进行比较。

颜色划分	Bh	ch	HI	L2
422	0.3802	0.3704	0.3321	0.3025
242	0.3975	0.3802	0.3444	0.3037
224	0.4346	0.4235	0.3852	0.3247

由实验结果可以看出,将蓝色空间划分为 4 份,将红色和绿色划分成

两份在各个距离度量下都会有很好的效果，当然，不排除是由于数据的特殊性导致此结果。

从划分的空间数来说，484 划分的效果明显好于 242 的效果，但是 484 划分会消耗更多时间。另外，也不一定是颜色空间划分地越细致最终的结果就会越好，因为我们要分类的物体多少在颜色上会有一些差异性，如果划分过于仔细就会放大这种差异性，适得其反。

### 4.3 错误划分分析

所有种类中分类效果最差的是 beach 和 elephant，在结果最好的 484Bh 里，elephants 的准确率仅仅为 0.15，beach 的准确率仅仅为 0.17。

elephant 辨别的准确率低是因为我们需要识别的三张 elephant 中有两张都是大象在草地中的图片，整个图片的主体为绿色，但是大象数据集中的绝大多图片的背景都是沙漠和土地，图片的主体色为黄色。与此相反的是，马的数据集中，绝大多数图片都为绿色背景（马在草地中奔跑），因此大象图片特别容易被识别为马。在这种情况下，仅仅依靠颜色无法识别不同的物体。可以引入图象区域划分技术，将原图首先划分为不同的区域，再对区域的重要性进行评估，在提取直方图的时候可以把区域重要性作为权重引入。

beach 的准确率低主要是因为 beach 数据集中有一部分只有沙滩没有大海，有些全是大海，还有些各占一半。尤其是全是沙滩的图片比较少，beach 测试集的第一图为 110.jpg，该图主要为沙滩，因此准确率非常低，第三张图 113.jpg 全是海洋，识别准确率稍微高一些。而且数据集中有些图片有人，有些图片没有人，人的肤色对我们这种依据颜色判断的方法也是一个很大的干扰。如果想要提高准确率的话，可以利用 CNN 对数据集中的图片提取特征，CNN 对于局部和全局特征的识别都比较适合，用在这里效果应该会比较好。

## 5 对原有方法的改进

### 5.1 引入距离 weight

我们提取的图片特征即为该图片的颜色直方图，颜色直方图只能表示该图的颜色特征，不能表示其位置特征，两张在视觉上差异很大的图片可能拥有同样的直方图。

通常我们可以认为，距离图片中心越近的像素越是重要，为此我们定义像素  $p(x,y)$  的 weight 为

$$weight(x,y) = \frac{1}{2\pi\sigma^2} e^{-((x-x_0)^2+(y-y_0)^2)/(2\sigma^2)} \quad (5)$$

上式  $(x_0, y_0)$  为图象中心坐标，假定图片的重要性符合正态分布。采用不同的  $\sigma$  可以控制正态分布的方差， $\sigma$  取得越大到中心点的距离对权重的影响越小。

取  $\sigma$  为 0.6 时，可以得到各方法组合的正确率

颜色划分	Bh	ch	HI	L2
242	0.4123	0.4037	0.3568	0.3148
484	0.4914	0.4877	0.4605	0.3926

可以看出，在引入了距离权重之后，各方法组合正确率提升约 1%，而且这种提升与各方法的选择几乎没有关系。从测试集的角度来说，beach 集合的准确率得到了约 4% 的提升，如我们之前所说，有些 beach 图片是一半大海一半沙滩，这种情况下图片中心要么是大海要么是沙滩，不管是哪一种得到权重上的强化，都会打破两种直方图混合带来的问题，从而带来颜色上的提升。其余大部分测试图片也有不同程度上的提升，个别图片的正确率有轻微的下降。

## 5.2 引入重要性 map

用到图片中心的距离来计算权重在有些情况下效果并不好，因为这张图片的关键因素可能并不在图片的中央，为了更科学地表示一张图片中各像素的重要性，我们引入 Boolean Map Saliency 技术来检测各像素点的重要性，BMS 能够较好得表示出图片中物体的拓扑结构，有利于我们根据物体的种类对图片进行划分。

图 1 表示 elephants/511.jpg 及其重要性灰度图,图 2 表示 elephants/511.jpg 及其重要性灰度图。

从图 1 和图 2 中可以看出，大象被亮度更高的值标出，从而说明大象的重要性在这两张图片当中比较高。在引入了重要性 map 之后，可以得到各方法组合的正确率

颜色划分	Bh	ch	HI	L2
242	0.3889	0.3827	0.3704	0.3630
484	0.4778	0.4778	0.4728	0.4432



(a) Small Box with a Long Caption



(b) Big Box

图 1: elephants/511.jpg 及其重要性灰度图



(a) Small Box with a Long Caption



(b) Big Box

图 2: elephants/511.jpg 及其重要性灰度图

将正确率表格与原始参数下的正确率表格进行对比可以看出，不同的距离度量方法对正确率的影响很大，Bh 度量下正确率反而下降了 1%，ch 下几乎没有什么变化，HI 下正确率提升了 2%，L2 下正确率提升了 5%，这主要是因为 BMP 技术除了标注出主要物体之外，往往还会标注出一些背景因素，因此直方图中除了主要物体的颜色信息外，往往还会混有一些背景物体的颜色信息，在 L2 度量下，对于这种多出来的背景物体信息不是很敏感，因为它是一种绝对度量，而不是一种相对度量，而其他几种方法则对这种“多出来”的信息更为敏感一些，所以会导致一些误判，所以 L2 会表现出比其他几种度量更加优异的特点。

当然，除了 BMP 之外还有其他重要性检测的方法，比如基于区域的检测方法可能会更好一些，但是由于本实验的主要任务是检索，所以对重要性检测方法的时间要求比较高，该方法必须要在短时间内作出重要性检测。

## 6 结论

从提高准确率的角度来说，可以将颜色集合划分得比较密集（过于密集也会影响准确率），采用 484 的划分方法可以说在时间和准确率上达到了一个折中。此外，若是采用 448 等将蓝色划分更密的方法会在本测试集上有很好的效果，但是在别的测试集上效果不一定会这么明显。引入距离权重和像素重要性也可以有效提升分类的准确率，但是距离权重对度量方法几乎没有要求，像素重要性权重对度量方法有特殊的要求，当然，可以采用一些改进的重要性检测技术来提升鉴别效果。