

目录

一、上机目的与要求	2
1.1 实验内容与目的	2
1.1.1 实验内容	2
1.1.2 实验目的	2
1.2 实验要求与评价准则	2
1.2.1 实验要求:	2
1.2.2 评价准则:	2
1.3 实验基本原理	3
1.3.1 基于内容的图像检索(CBIR)	3
1.3.2 LBP(Local Binary Pattern)算法	4
1.3.3 SIFT 算法 & BoW	5
1.3.4 颜色矩特征	5
1.3.5 评价指标: AP、mAP	6
二、上机内容	7
2.1 方法与步骤	7
2.1.1 数据集	7
2.1.2 基于 LBP 特征的图像检索方法	8
2.1.3 基于 SIFT(BoW)特征的图像检索方法	9
2.2 结果与分析	12
2.2.1 基于 LBP 特征、颜色矩特征的图像检索结果	12
2.2.1 基于 SIFT(BoW)特征、颜色矩特征的图像检索结果	13
三、总结	16

一、上机目的与要求

1.1 实验内容与目的

1.1.1 实验内容

在所给电商数据集上测试基于 LBP 特征的图像检索；实现基于 SIFT(BoW)特征的图像检索；在此基础上，以 AP/MAP 为评价指标，测试不同特征提取方式对应的图像检索算法的性能。

1.1.2 实验目的

1. 了解并掌握 LBP、SIFT(BoW)、颜色矩的特征提取算法；
2. 掌握信息检索常用评价指标 P、R、AP/MAP 的意义、计算方法与代码实现；
3. 掌握基于内容的图像检索（CBIR）算法的基本流程

1.2 实验要求与评价准则

1.2.1 实验要求：

1. 测试不同的特征提取方式：
LBP；
SIFT(BoW)；
2. 特征相似性度量方式：
欧氏距离；
3. 可视化查询结果；
可视化编号为 1,201,401,601,801,1001,1201,1401,1601,1801 图像作为查询的 Top 12 个检索结果（检索对象可包含查询图像）。

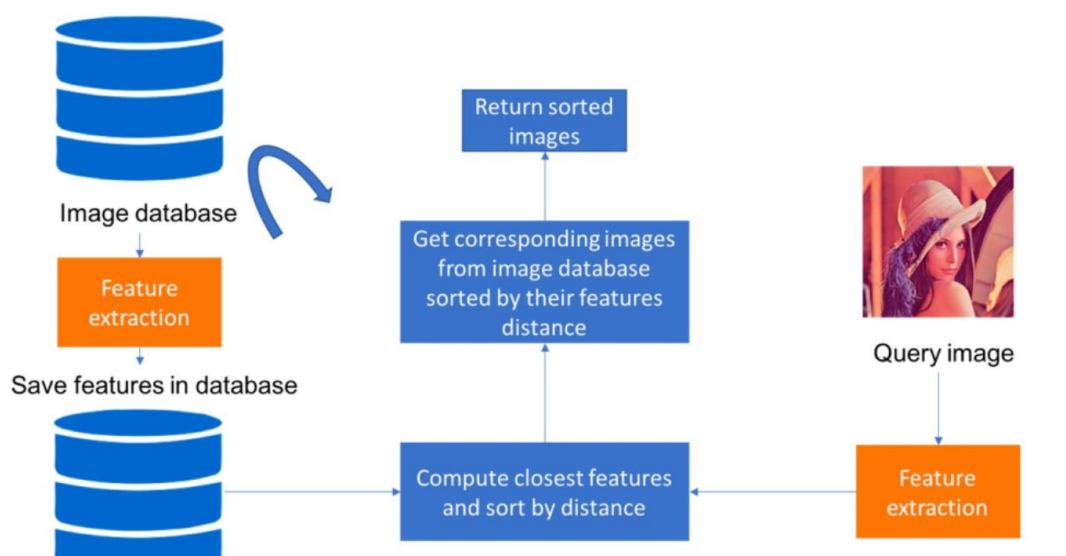
1.2.2 评价准则：

计算 query 图像（每个类别的第一幅图像，编号为 1,201,401,601,801,1001,1201,1401,1601,1801）在数据集中的检索结果（Top 12），计算 mAP 值。

1.3 实验基本原理

1.3.1 基于内容的图像检索(CBIR)

基于文本的图像检索方法始于上世纪 70 年代，它利用文本标注的方式对图像中的内容进行描述，从而为每幅图像形成描述这幅图像内容的关键词，比如图像中的物体。这种基于文本描述的方式所带来的缺陷也是非常明显的:首先在大规模图像数据上要完成这一过程需要耗费大量的人力与财力;其次，用户有时很难用简短的关键词来描述出自己真正想要获取的图像;最后，人工标注过程不可避免的会受到标注者的认知水平、言语使用以及主观判断等的影响，因此会造成文字描述图片的差异。



CBIR 本质也是"召回->排序"的框架，主要流程如下:

1. 输入待检索图像
2. 特征提取。抽取高效稳定可重复的特征提取(比如 LBP,SIFT,CNN 等)
3. 对图像数据库建立图像特征提取与索引存储
4. 抽取检索(Query)图像特征，构建特征向量
5. 设计检索模块，包含相似性度量准则，排序，搜索等.
6. 返回相似性较高的结果.

在本次实验中，图像特征的提取分别采用了 LBP、LBP+颜色矩、SIFT 和 SIFT+颜色矩的特征提取方法；相似度的计算采用欧氏距离为衡量指标；采用 AP/MAP 作为模型性能的最终评价指标。

图像检索所面临的挑战:

- 图像光照变化
- 尺度变化
- 视角变化
- 遮挡
- 背景混乱
- 仿射变换

1.3.2 LBP(Local Binary Pattern)算法

LBP 是一种描述图像特征像素点与各个像素点之间的灰度关系的局部特征的非参数算法，同时也是一张高效的纹理描述算法。

纹理是物体表面的自然特性，它描述图像像素点与图像领域之间的灰度空间的分布关系，不会因为光照强弱而改变图像的视觉变化。

LBP 算法首次提出于 1994 年，主要是使用 8 邻域位置的局部关系，具有灰度不变性；随后在 2002 年提出了其改进版《Multiresolution gray-scale and rotation invariant texture classification with local binary patterns》主要引入了圆形模式、旋转不变性、和等价模式。

下图为一个 LBP 算法的计算例子：

90	74	63
103	85	134
99	83	91

1	0	0
1		1
1	0	1

1	0	0
128		8
64	0	16

该纹理单元的中心灰度值(阈值)为 85，其余 8 个相邻的灰度值分别为：90、74、63、134、91、83、99、103、90。根据这个阈值计算的定理进行计算，通过阈值计算后，根据图中箭头所指的方向可以得到一个局部二进制的模式，即：LBP 模式=1101001，且 LBP=217。

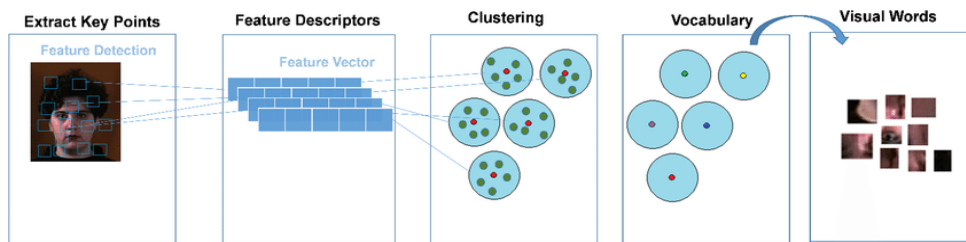
通过以上 LBP 算法的计算例子可以看出 LBP 值仅由中心像素点灰度值和相邻八个像素点的灰度值决定，LBP 特征即表示了其中心像素点的特征，其值的计算还需要和周围的像素点计算完成，所以，LBP 特征与周围有很大的关系，它包含了图像点，又包含了边缘和局部的特征分布信息。

1.3.3 SIFT 算法 & BoW

SIFT 算法核心是将一幅图像用描述子表示，这些特征点具有尺度不变性，就相当于图像本身，但是相对于图像文件本身，在计算机中处理起来方便的多了。详细的 SIFT 实现过程分为如下四个步骤：

- 1、检测尺度空间的极值点；
- 2、抽取稳定的关键点；
- 3、为每个关键点指定一个或者多个方向；
- 4、生成特征点描述子。

BoW 其大概过程：首先提取图像集特征的集合，然后通过聚类的方法聚出若干类，将这些类作为 **codebook**，即相当于 **words**，最后每个图像统计 **codebook** 中 **words** 出现的频数作为输出向量，就可以用于后续的分类、检索等操作。



1.3.4 颜色矩特征

颜色是彩色图像最重要的内容之一，被广泛用于图像检索中。但从图像中提取颜色特征时，很多算法都先要对图像进行量化处理。量化处理容易导致误检，并且产生的图像特征维数较高，不利于检索。

AMA Stricker 和 M Orengo 提出了颜色矩的方法，颜色矩是一种简单有效的颜色特征表示方法，有一阶矩(均值,mean)、二阶矩(方差, variance)和三阶矩(斜度,skewness)等，由于颜色信息主要分布于低阶矩中，所以用一阶矩，二阶矩和三阶矩足以表达图像的颜色分布，颜色矩已证明可有效地表示图像中的颜色分布，该方法的优点在于：不需要颜色空间量化，特征向量维数低；但实验发现该方法的检索效率比较低，因而在实际应用中往往用来过滤图像以缩小检索范围。

三个颜色矩的数学定义如下：

$$\mu_i = \frac{1}{N} \sum_{j=1}^N p_{i,j}$$

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{i,j} - \mu_i)^2 \right)^{\frac{1}{2}}$$

$$s_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{i,j} - \mu_i)^3 \right)^{\frac{1}{3}}$$

其中， $p_{i,j}$ 表示彩色图像第 j 个像素的第 i 个颜色分量， N 表示图像中的像素个数。（即为 RGB 通道值）

1.3.5 评价指标：AP、mAP

AP (Average precision) 是主流的目标检测模型的评价指标。这一指标同时考虑了 Precision 和 Recall，MAP 就是不同类别 AP 的平均值。

AP 就是平均精准度，简单来说就是对 Precision 求均值。AP 的计算有 11 点法和 P-R 曲线下求面积积分的方法。

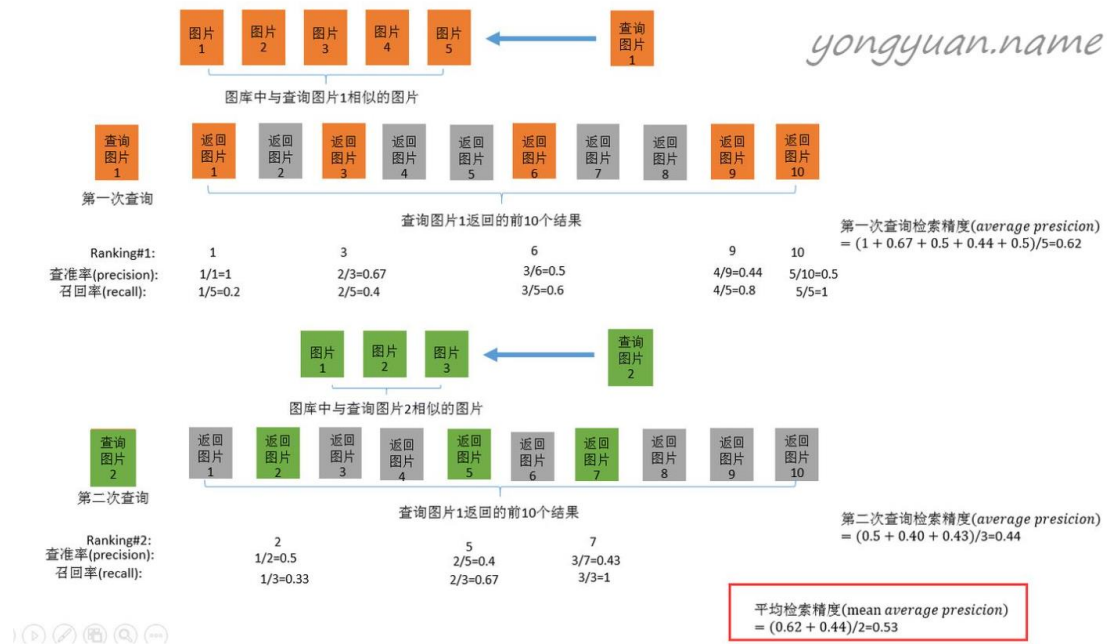
在本实验中，采用 11 点法的近似方法。选取当 Recall $\geq 0, 0.1, 0.2, \dots, 1$ 共 11 个点时的 Precision 最大值，AP 是这 11 个 Precision 的平均值，此时只由 11 个点去近似 PR 曲线下面积。

公式如下：

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \dots, 1\}} P_{interp}(r)$$

$$mAP = \frac{\sum_{q=1}^Q AP(q)}{Q}$$

示例：



相关代码实现如下：（详见 main.m）

```
% 计算topnum中与查询图像类别一致的样本数目
topidx = index(1:topnum+1);
number = 0;
precisions = 0;
gnd = [];
for i=2:topnum+1
    if fix(topidx(i)/200) == fix(topidx(1)/200)
        number = number + 1;
        gnd = [gnd, i];
        precisions = precisions + number / (i-1);
    end
end

% count = count + 1;
% count
ap = 0;
if number ~= 0
    ap = ap + precisions / number
    % avg_ap = avg_ap + precisions / number;
end
end
```

二、上机内容

2.1 方法与步骤

2.1.1 数据集

使用电商数据集 image，存放在 ./CBIR/image 中。

数据集一共 10 类，每类共 200 幅图片。编号为 1-2000.jpg。

2.1.2 基于 LBP 特征的图像检索方法

LBP 特征提取的代码已经在 ./CBIR/lbp.m 中实现。

预先提取的 LBP 特征位于 ./CBIR/statxture.mat 中。只需直接在 ./CBIR./main.m 中读取预先提取的 LBP 特征，然后进行后续计算即可。

```
1 - S = load('statxture.mat');
2 - C = load('colorMom.mat');
3 - SIFT = load('bof_feature.mat');
4
5 - StatxtureData = S.Istatxture;
6 - ColorData = C.IcolorMom;
7 - SIFTData = SIFT.bof_train;
```

LBP 是一种纹理描述算法，不包含颜色特征。所以在实际使用中，颜色矩特征常常与 LBP 特征一起使用来提高检索的正确率。因此，分别使用 LBP 特征和 LBP + 颜色矩特征来计算相似度。

LBP 相似度计算如下：

LBP:

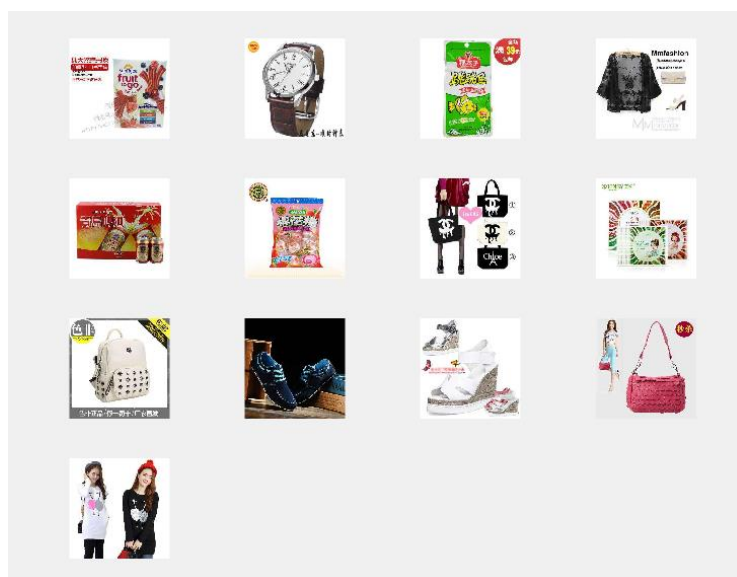
```
order(1,i)=CalDis(StatxtureData (:, RetrieveImg), StatxtureData (:, i));
```

LBP + 颜色矩:

```
order(1,i)=CalDis(ColorData(:,RetrieveImg),ColorData(:,i))*0.2+CalDis(StatxtureData (:,RetrieveImg), StatxtureData (:,i))/1000000*0.8;
```

可视化检索效果如下：

(以 201 检索结果为例)



2.1.3 基于 SIFT(BoW)特征的图像检索方法

基于实验一 BoW 的基础之上，提取图像的 SIFT 特征；并使用 K-means 对提取特征聚类，生成 codebook；最后依据图像中每个特征距离 codebook 中特征的最近距离，给所有的图像的 SIFT 特征定量，得到图像的特征表示。

生成图像 SIFT(BoW)特征的部分位于 ./SIFT； ./img/CBIR 存放图像数据集， ./divide 为数据集预处理部分， ./matlab 为特征表示生成部分

1. 首先需要将待检索图片整理为按类别分放的格式，方便使用：

运行 ./divide/divide_image.py 将图像按分类放置：

```
def move_picture(pid, dir_id):  
    old_path = pid + ".jpg"  
    new_path = "./" + dir_id  
    shutil.move(old_path, new_path)  
    return
```

电商数据集中，连续的 200 张图像为同一类别，共整理为 10 类。

0	29/05/2021 17:07
1	29/05/2021 15:24
2	29/05/2021 15:26
3	29/05/2021 15:28
4	29/05/2021 15:31
5	29/05/2021 15:35
6	29/05/2021 15:36
7	29/05/2021 15:38
8	29/05/2021 15:40
9	29/05/2021 15:22

2. SIFT(BOW)特征向量的提取在 ./SIFT/matlab/main.m 实现
dataset_dir 为 CBIR

```
28 % DATASET  
29 - dataset_dir='CBIR';  
30
```

将训练集大小设为每类图像的数量(200)，codebook 维数为 500

```
% number of images selected for training (e.g. 30 for Caltech-101)  
num_train_img = 200;  
% number of images selected for test (e.g. 50 for Caltech-101)  
num_test_img = 0;  
% number of codewords (i.e. K for the k-means algorithm)  
nwords_codebook = 500;
```

需要注意的是，在生成数据集并进行划分的函数 `create_dataset_split_structure.m` 中，须使用按自然数顺序的排序方式替代随机排序的方式。这是因为原始实现中会随机的打乱样本的顺序重新划分数据集，而在本次实验中提取到的特征顺序和图像顺序需要从 1-2000 的排列顺序一致。按自然数顺序排序的代码 `sort_nat` 需要在 matlab 官网上下载使用：
https://ww2.mathworks.cn/matlabcentral/fileexchange/10959-sort_nat-natural-order-sort

```
% 按自然数顺序的排序方式替代随机排序
% 提取到的特征的顺序和图像从 1-2000 的排列顺序一致
% ids = randperm(length(imgdir));
imgdir = sort_nat({imgdir.name});
ids = 1:1:(length(imgdir));
```

如果不进行此排序，则 matlab 默认的读取顺序是按照二进制数读取，会导致特征向量和图片的顺序不匹配。

数据集划分结果储存在 `./SIFT/img/CBIR/split.mat` 中；

1x10 struct with 5 fields

Fields	n_images	classname	files	train_id	test_id
1	200	'0'	1x200 cell	1x200 logical	1x200 logical
2	200	'1'	1x200 cell	1x200 logical	1x200 logical
3	200	'2'	1x200 cell	1x200 logical	1x200 logical
4	200	'3'	1x200 cell	1x200 logical	1x200 logical
5	200	'4'	1x200 cell	1x200 logical	1x200 logical
6	200	'5'	1x200 cell	1x200 logical	1x200 logical
7	200	'6'	1x200 cell	1x200 logical	1x200 logical
8	200	'7'	1x200 cell	1x200 logical	1x200 logical
9	200	'8'	1x200 cell	1x200 logical	1x200 logical
10	200	'9'	1x200 cell	1x200 logical	1x200 logical

运行 `./SIFT/matlab/main.m`，其中将特征结果保存在本地的命令是：
`save bof_feature bof_train;`

最终的特征结果储存在 `./SIFT/matlab/bof_feature.mat` 中，是 2000*500 维的矩阵，每个图像对应一个 500 维的 SIFT 特征表示；将其转置为 500*2000，与下一步统一格式。

将 `bof_feature.mat` 复制到 `./CBIR` 目录之下，并在 `./CBIR./main.m` 中读取 SIFT 特征：

```

1- S = load('statxture.mat');
2- C = load('colorMom.mat');
3- SIFT = load('bof_feature.mat');
4
5- StatxtureData = S.Istatxture;
6- ColorData = C.IcolorMom;
7- SIFTData = SIFT.bof_train;

```

类比于 LBP，SIFT 相似度计算如下：

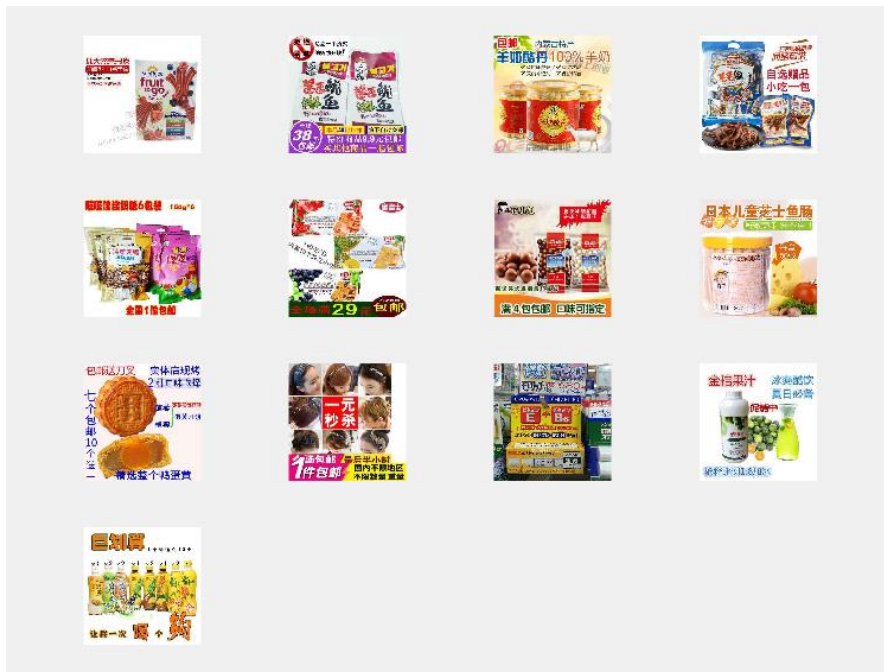
SIFT:

$\text{order}(1,i)=\text{CalDis}(\text{SIFTData}(:, \text{RetrieveImg}), \text{SIFTData}(:, i));$

SIFT + 颜色矩:

$\text{order}(1,i)=\text{CalDis}(\text{ColorData}(:, \text{RetrieveImg}), \text{ColorData}(:, i)) * 0.2 + \text{CalDis}(\text{SIFTData}(:, \text{RetrieveImg}), \text{SIFTData}(:, i)) / 1000000 * 0.8;$

可视化检索效果如下：（以 201 检索结果为例）



2.2 结果与分析

2.2.1 基于 LBP 特征、颜色矩特征的图像检索结果

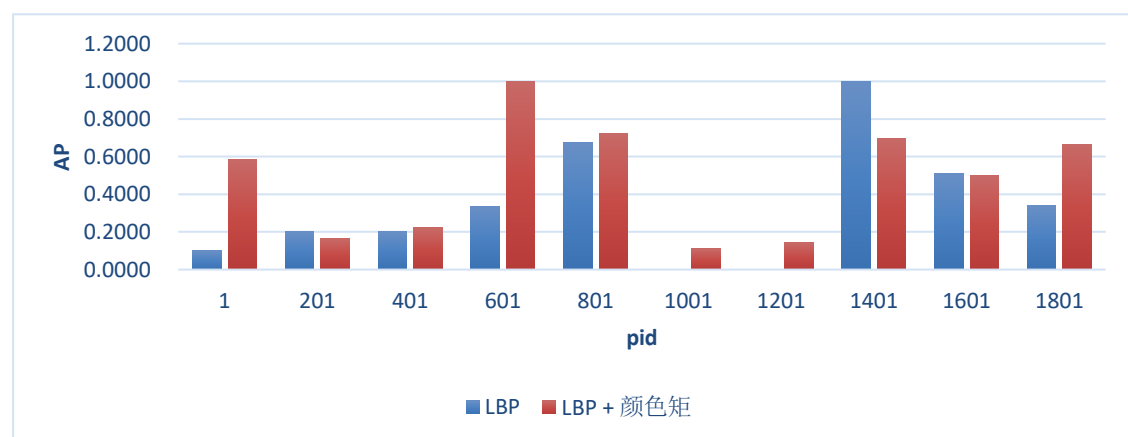
待检索图像序列：

[1,201,401,601,801,1001,1201,1401,1601,1801]

每个类别选取第一张图片

实验数据如下：

LBP										
Classes	0-Dress	1-Food	2-Pants	3-Bag	4-Furniture	5-Bottle	6-Maquillage	7-Shoes	8-Jewelry	9-Jacket
pid	1	201	401	601	801	1001	1201	1401	1601	1801
AP	0.1000	0.2000	0.2000	0.3333	0.6772	0.0000	0.0000	1.0000	0.5105	0.3409
MAP	0.3362									
LBP + 颜色矩										
Classes	0-Dress	1-Food	2-Pants	3-Bag	4-Furniture	5-Bottle	6-Maquillage	7-Shoes	8-Jewelry	9-Jacket
pid	1	201	401	601	801	1001	1201	1401	1601	1801
AP	0.5833	0.1667	0.2222	1.0000	0.7236	0.1111	0.1429	0.6984	0.5000	0.6667
MAP	0.4815									



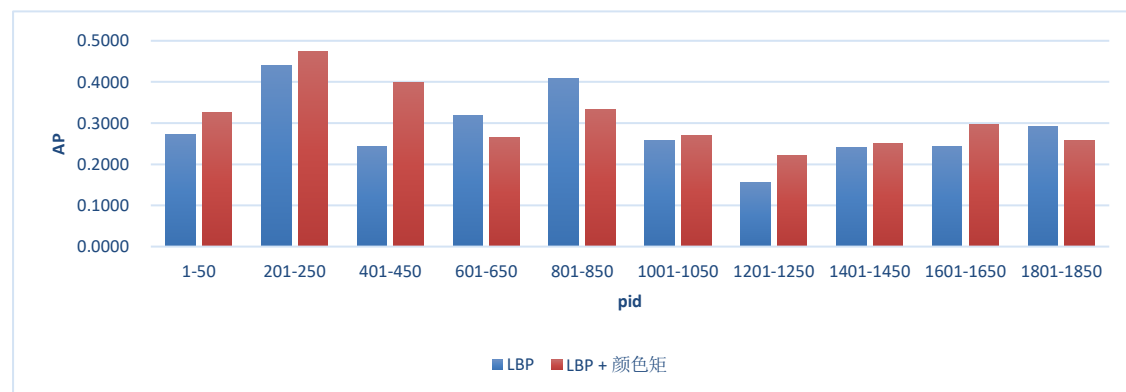
待检索图像序列：

[1:50,201:250,401:450,601:650,801:850,1001:1050,1201:1250,1401:1450,1601:1650,1801:1850]

每个类别选取前 50 张图片

实验数据如下：

LBP										
Classes	0-Dress	1-Food	2-Pants	3-Bag	4-Furniture	5-Bottle	6-Maquillage	7-Shoes	8-Jewelry	9-Jacket
pid	1-50	201-250	401-450	601-650	801-850	1001-1050	1201-1250	1401-1450	1601-1650	1801-1850
AP	0.2714	0.4400	0.2431	0.3191	0.4080	0.2585	0.1557	0.2418	0.2426	0.2925
MAP	0.2873									
LBP + 颜色矩										
Classes	0-Dress	1-Food	2-Pants	3-Bag	4-Furniture	5-Bottle	6-Maquillage	7-Shoes	8-Jewelry	9-Jacket
pid	1-50	201-250	401-450	601-650	801-850	1001-1050	1201-1250	1401-1450	1601-1650	1801-1850
AP	0.3256	0.4749	0.3980	0.2655	0.3341	0.2704	0.2220	0.2502	0.2963	0.2578
MAP	0.3095									



由上述数据可知：

1. LBP+颜色矩特征的检索效果强于单独的 LBP 特征，对于 Pants 类的作用效果最为明显；
2. 同一类别不同图像的检索效果差距较大，导致 50 张图像的平均 MAP 与 1 号图像有 0.1 左右的差距；
3. LBP 特征对于 Food 与 Furniture 类的检索效果较好，对于 Maquillage 类效果较差。

2.2.1 基于 SIFT(BoW)特征、颜色矩特征的图像检索结果

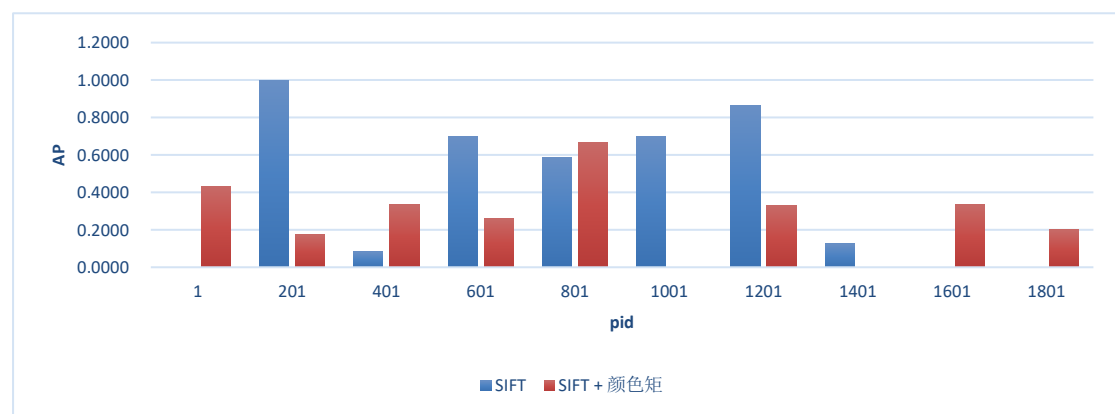
待检索图像序列：

[1,201,401,601,801,1001,1201,1401,1601,1801]

每个类别选取第一张图片

实验数据如下：

SIFT										
Classes	0-Dress	1-Food	2-Pants	3-Bag	4-Furniture	5-Bottle	6-Maquillage	7-Shoes	8-Jewelry	9-Jacket
pid	1	201	401	601	801	1001	1201	1401	1601	1801
AP	0.0000	1.0000	0.0833	0.7000	0.5868	0.7000	0.8667	0.1250	0.0000	0.0000
MAP	0.4062									
SIFT + 颜色矩										
Classes	0-Dress	1-Food	2-Pants	3-Bag	4-Furniture	5-Bottle	6-Maquillage	7-Shoes	8-Jewelry	9-Jacket
pid	1	201	401	601	801	1001	1201	1401	1601	1801
AP	0.4333	0.1742	0.3333	0.2611	0.6667	0.0000	0.3321	0.0000	0.3333	0.2000
MAP	0.2734									



待检索图像序列：

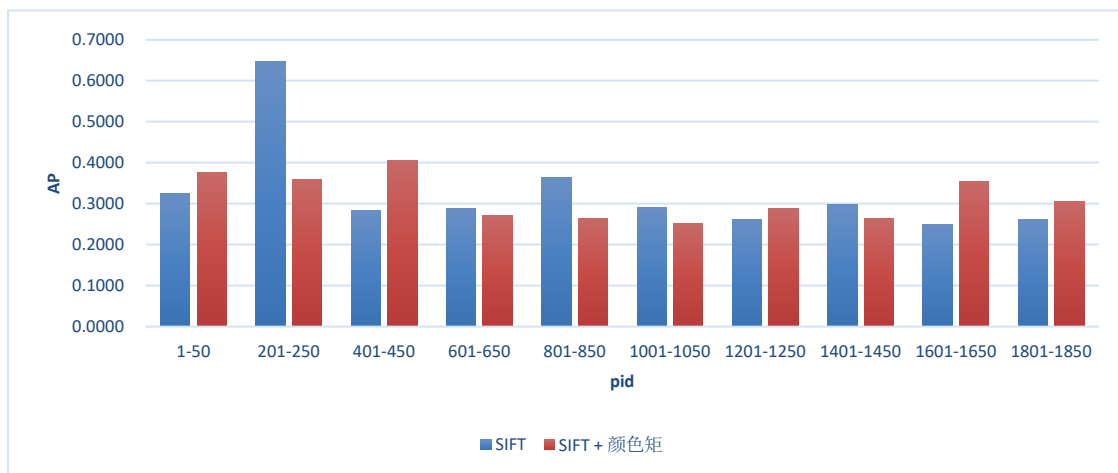
[1:50,201:250,401:450,601:650,801:850,1001:1050,1201:1250,1401:1450,1601:1650,1801:1850]

每个类别选取前 50 张图片

实验数据如下：

SIFT										
Classes	0-Dress	1-Food	2-Pants	3-Bag	4-Furniture	5-Bottle	6-Maquillage	7-Shoes	8-Jewelry	9-Jacket
pid	1-50	201-250	401-450	601-650	801-850	1001-1050	1201-1250	1401-1450	1601-1650	1801-1850
AP	0.3256	0.6482	0.2826	0.2875	0.3635	0.2918	0.2613	0.2982	0.2501	0.2614
MAP	0.3270									

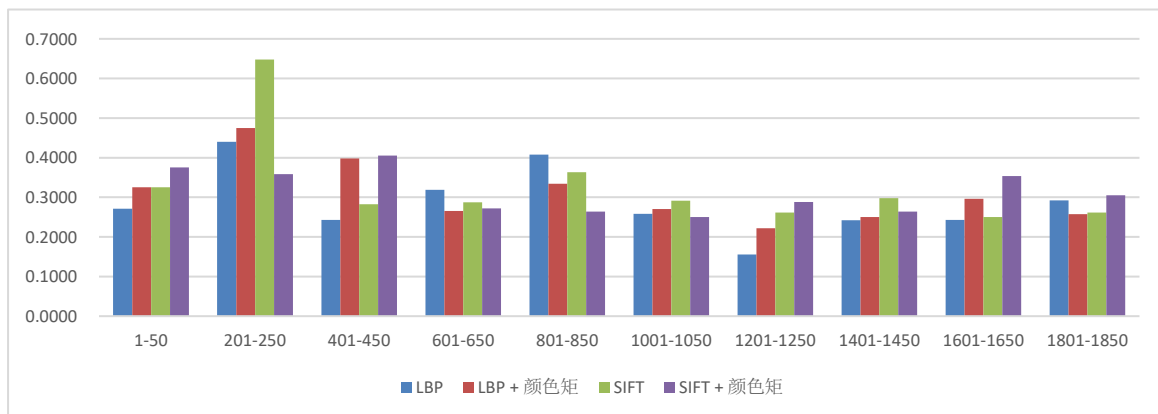
SIFT + 颜色矩										
Classes	0-Dress	1-Food	2-Pants	3-Bag	4-Furniture	5-Bottle	6-Maquillage	7-Shoes	8-Jewelry	9-Jacket
pid	1-50	201-250	401-450	601-650	801-850	1001-1050	1201-1250	1401-1450	1601-1650	1801-1850
AP	0.3758	0.3583	0.4053	0.2718	0.2636	0.2505	0.2881	0.2638	0.3540	0.3054
MAP	0.3137									



由上述数据可知：

1. SIFT+颜色矩特征的检索效果未明显强于单独的 SIFT 特征，仅对其中的 5 类图像的检索效果有提升作用，且对于 Food 类的起到了明显的反作用；
2. 同一类别不同图像的检索效果差距较大，导致 50 张图像的平均 MAP 与 1 号图像有 0.1 左右的差距；
3. SIFT 特征对于 Food 与 Furniture 类的检索效果较好，对于 Maquillage 和 Jewelry 类的效果较差。

综合实验数据，对比如下：



由图可见：

1. 引入颜色矩特征对 LBP 特征的图像检索效果有显著帮助，但对于 SIFT 特征来说，效果并不十分明显；
2. 检索效果排序大致为 $LBP < LBP + \text{颜色矩} < SIFT + \text{颜色矩} < SIFT$ ；
3. 从类别来看，四种方式组合对于 Food 类别检索效果最佳，对 Maquillage 类效果较差。

三、总结

通过本次上机实验，我掌握了基于内容的图像检索(CBIR)方法。在实验过程中，我分别抽取了电商数据集中图像的 LBP、SIFT 与颜色矩特征；使用其生成图像的对应的特征向量；赋予不同特征不同的权重，使用欧氏距离计算其相似度；统计 TOP12 中检索正确的结果；以 AP/MAP 为评价指标，测试不同特征提取方式对应的图像检索算法的性能。

在掌握了 CBIR 基本方法与流程之后，我依据老师的实验要求，分别测试了基于 LBP 与 SIFT 特征的图像检索的效果，并结合颜色矩进行了组合实验。记录实验数据，绘制数据表格与图像，分析相应结论。

在这次实验中，我也深刻的了解了基于内容的图像检索方法相较于基于文本的图像检索的显著优势：降低主观性偏差，减少时间成本，更加充分的表达图像中包含的丰富的信息。