

Image Retrieve System Based on ANN

张佳瑜
2012013304
软件 21

zhangjiayu55@163.com

王程鹏
2012013309
软件 22

mhnfwln@126.com

杨妍喆
2012013323
软件 22

yangyanzhe@126.com

摘要

在这篇报告中，我们将主要描述本图像检索系统的设计与核心思想，介绍在过程中所使用的技术和做的尝试。同时，还有我们得到的结论和思考。

关键词

SIFT, 感知哈希, 小波变换, 人工神经网络

一、前言

随着多媒体技术和互联网技术的发展，基于内容的图像检索吸引了越来越多的学者加入到它的研究中来。如何快速而准确地搜索到数据库中类似的图片，成为一个挑战而有趣的课题。

本图像检索系统分为两部分，一部分为图像特征的提取，另一部分为人工神经网络的搭建。特征提取方面，基于项目一的基础，我们尝试了更多的特征提取方式，希望可以从颜色、纹理、形状等多个方面提取特征，用于人工神经网络的学习。采用人工神经网络，则可以很好地学习各个特征的权重，提取出高层特征。程序在ubuntu 12.04 下运行、测试。

二、图像特征的提取

1.HSV

HSV 颜色直方图的介绍在此不再赘述。实现方面，我们没有采用 H、S、V 三维空间，而是将 H、S、V 分别作为单一坐标，提取落在 H 不同区间、S 不同区间、V 不同区间的统计，从而降低了维度。在单位的划分上，我们发现 H、S、V 维数为 2:1:1 时有较好结果。

特征提取之后，我们还进行了归一化处理，得到每个区间内的统计量的百分比，这样的处理一方面方便提升人工神经网络的性能，另一方面，也更加有价值，避免了图片大小带来的干扰。

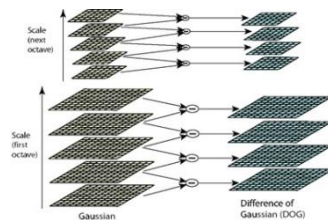
2. SIFT 特征

2.1 SIFT 特征简介

SIFT 特征是一种检测局部特征的算法，算法流程如下：

首先，对原图作高斯卷积，得到新的图片，并新得到的图片依次做不同参数下的高斯卷积，最终构建出的高斯金字塔如下所示。

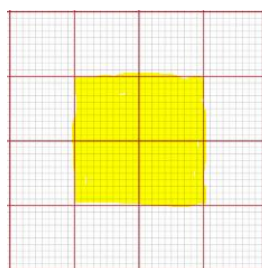
接着，选取同一组中相邻的图片相减，得到高斯金字塔的差商金字塔 (Difference of Gaussian)。在差商金字塔的每层图片中提取出极值点，经过进一步的筛选后得到特征点的集合。最后对每个特征点计算梯度方向，利用关键点邻域像素的梯度方向分布特性为每个关键点指定方向参数，使算子具备旋转不变性，最终每个特征点对应 128 维的特征向量。在本系统中，我们利用了 OpenCV 中的 Sift 函数提取了特征。



2.2 SIFT 特征的分析与改进

单纯地从特征提取的角度上看，SIFT 特征不只具有尺度不变性，即使改变旋转角度，图像亮度或拍摄视角，仍然能够得到好的检测效果。但是在测试的过程中发现其存在的缺点有：特征点个数不定，需要做进一步处理得到固定维数的特征向量；由于 Sift 算子提取特征较为精细，提取特征点较多的图片时效率不高。

针对第一个问题，我们采用了分块处理的方法，提取出位于图片中心位置（下图中黄色区域）的特征点，并计算四块中特征点横纵坐标、梯度向量夹角的均值与方差，得到 24 维的特征向量。



之后我们将特征的横纵坐标除以图片的 row 和 col，采用相对坐标可以提高找到缩放图的概率。针对第二个问题，我们进一步尝试了 SURF 特征。

2.3 SURF 特征

Surf 特征与 Sift 特征提取过程类似，构建尺度空间模型后在差分金字塔中提取出特征点。Surf 算子用 box_filter 代替了高斯滤波器，提取特征的平均时间开销比 Sift 特征小。但 Surf 提取时没有 Sift 特征精细，尺度、旋转不变性不如 Sift 特征。用人工神经网络测出在做类别判断时 Surf 特征与 SIFT 特征结果相近。

3. 感知哈希

3.1 感知哈希简介

类似于 key-to-address 的思想，均值 Hash 算法是将图片分割成 8*8 的方格，选取某种方式（如取 RGB 的平均值）将方格缩成一个像素点，得到缩小后得图片。再计算新得到图片的平均值，若像素点的值大于平均值则赋值为 1，否则赋值为 0。最终得到长度为 64 且由 0、1 组成的特征向量。

均值 Hash 的算法较为朴素，pHash 算法在前者基础上做出改进。图片的内容大多由低频成分反映，高频成分如区域边界、噪声点往往不能准确反映其内容。pHash 特征提取的思路为：将图片缩小成 32*32 的图片，储存在 32*32 维的矩阵中。对缩小后的图片做 DCT（离散余弦变换），变换后的矩阵左上角的频率成分较低，故选取左上角的 8*8 的 64 个像素点作为原图对应的特征图片。

3.2 算法分析与实现



pHash 算法具有较好的伸缩不变性。例如：左图在 1024*768，316*237 两种情况下提取的特征分别为

(130,255,163,220,31,0,110,
112,129,135,198,136,131,1

06,89,159,133,117,207,124,132,124,95,144,120,121,160,12
4,136,109,125,141,117,143,138,132,131,109,131,136)

(128,193,141,255,34,0,122,136,119,143,199,145,139,115,
88,168,121,123,214,128,127,117,97,152,101,121,151,124,1
39,102,124,139,102,124,137,110,154,132,126,134,109,130,
137)

可以看出，所提取的特征结果是非常相近的。

4 Gabor 小波变换

4.1 Gabor Wavelet 简介

Gabor 小波与人类视觉系统中的视觉刺激响应非常相似，在提取图片的频率域信息方面具有良好的性质。因此，Gabor 变换也经常被称为数学显微镜，常被用来进行纹理特征的提取。Gabor 函数是一个用高斯函数调制的负正弦函数，能够在给定区域内提取出局部的频域特征。

$$g(x,y;\lambda,\theta,\varphi,\sigma,\gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi \frac{x'}{\alpha} + \varphi\right)\right)$$

4.2 算法的实现与分析

Gabor 变换在查找相关资料后，使用了开源的 cvGabor.h 与 cvGabor.cpp，可以生成相关方向和尺度的 gabor，显示其实部、虚部或模。

调用 gabor 函数，设置参数得到 8 个方向，4 个尺度上的 32 张滤波变换图片，分别对滤波结果提取其均值和方差作为特征向量。为了方便人工神经网络进行学习，将向量值进行了归一化和放大处理，使得特征值为 (0, 1) 区间非聚集分布。

通过该单一特征量测量准确度，准确度略低于 30%。其可能的原因在于参数的选取使得特征表现地并不明显，使得差异性比较小。在测试中，它对图片的背景信息的区分并不好，如右图。



三、人工神经网络技术

3.1 人工神经网络简介

神经网络部分程序由训练部分与运行部分组成。fann.h 中包含了训练部分的 c++ 代码。同时，为了对不同特征在人工神经网络下的性能有预先的了解，我们用 python 与 fann2.1 写了很多神经网络训练与测试的代码。这些代码可以在 python 文件夹下找到。神经网络运行的代码包含在主程序中。

3.2 图片检索系统测试结果

对于单个特征的测试，结果如下表所示

H	S	V	准确率
8	8	0	35.2%
6	6	6	36.8%
12	6	6	38.1%

Fig. 1. HSV 直方图取不同划分的准确率

特征种类	准确率
Surf	34.5%
Wavelet	23.9%
pHash	~20%
HSV	>=40%

Fig. 2. 不同特征的单一测量准确率

对于 hsv, phash, color_feature, surf, 小波特征的混合测试正确率为 56.2%。在系统中，我们采用

(H,S,V) = (8,8,0), (6,6,6), (12,6,6)、surfFeature，测试结果为 49.5%。由于 pHash 特征在单一测量下正确率较低，我们最终选择将 HSV, Surf, Wavelet 组合成特征向量。

3.3 分析与讨论

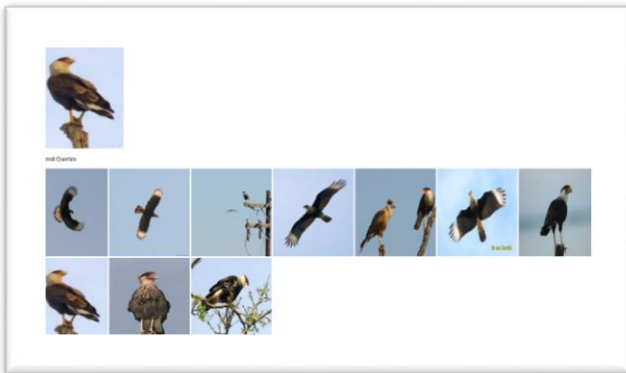
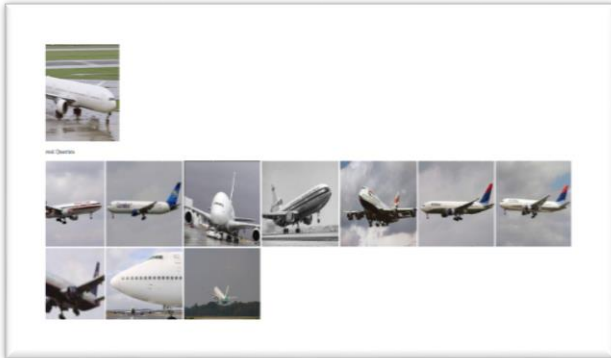
首先，数据的归一化对神经网络的训练很有影响。如果数据不做归一化，虽然理想状态神经网络第一层应当收敛到比较小的值，但实际测试发现神经网络很难收敛，隐藏层节点作用被削弱了，最终数据只

是隐藏层节点输出的线性组合，从而影响测试效果。

此外，我们还研究了不同隐藏层节点个数对神经网络正确率的影响。节点数目从10个开始逐渐增加时，神经网络正确率逐步提高。但当节点数达100左右时，神经网络正确率不再增加。考虑到过拟合等问题，我们建议隐藏层节点数为100-120。

四、测试结果

选取飞机、鸟类图片搜索，结果如下所示，搜得的十张图片与输入图片类别相同。



五、结论与讨论

人工神经网络在图片分类方面有很大的优势，尤其在处理高维特征时，与其他决策机制如随机森林相比成功率更高，并且在 Linux 系统下时间性能优越。但是基于类别的图片检索存在一个难以解决的问题：若在经过特征提取和神经网络后输出的图片类别有误，之后基于 HSV 特征做 k 近邻查找得到的结果均不为期望得到的结果。因此依赖类别判定的图片检索容易只出现全对、全错两种情况。

此外我们认为限制人工神经网络分类性能因素的不是网络本身，而是特征提取过程有严重的信息丢失。对于错误的分类，从特征的角度看，它们与被分至的图片类有很多相似之处。新的特征的加入不会导致维度灾难，而会提升网络的性能。对图像进行分块，统计每一块的

HSV 值可以保留更多的图像语义信息，加入这一特征有可能获得更好的效果。

我们还尝试了用 SVM 做类别判断。但是由于时间原因，没有取得很好的效果。

六、致谢

感谢靳老师和助教在项目完成中给予的帮助和支持。在图像检索系统的构建中，我们获益匪浅。

在构建系统的过程中，我们尝试了很多不同的技术，除了 Hu 矩阵、HSV，还有上述提出的 SIFT、SURF、pHash、Wavelet 四种方式，此外，还尝试了区域分割、高斯平滑、参数调节等多种优化方式。在建立系统的过程中，培养了有效搜索信息、阅读文献与独立解决问题的能力。虽然有些尝试失败了，但是这些工作都是进一步更好地完成的基础。而不断地总结分析、改进方案，本身就是一种进步和提高。

在项目中，特征提取与人工神经网络都做了很多的工作，每个特征都有 300 多行自己写的代码。合并时我们进行了代码重构，最终实现了高效地检索，并给出了简洁的工程。

七、参考文献与库的使用

- [1] David, G. Lowe. Object Recognition from Local Scale-Invariant Features [D]. Vancouver, B.C., V6T 1Z4, Canada: David G. Lowe, 1999.
- [2] Sift 特征提取分析
<http://blog.csdn.net/abcjennifer/article/details/7639681>
- [3] 局部特征三——SURF 特征总结
http://blog.csdn.net/jwh_bupt/article/details/6564835
- [4] 相似图片搜索的三种哈希算法
<http://blog.csdn.net/zmazon/article/details/8618775>
- [5] Gabor 变换中使用的 cvGabor.cpp 与 cvGabor.h 来自
http://download.csdn.net/detail/yao_zhuang/490114
- [6] David Barina, Gabor Wavelets In Image Processing
<http://www.feec.vutbr.cz/EEICT/2011/sbornik/03-Doktorske%20projekty/09-Grafika%20a%20multimedia/01-xbarin02.pdf>
- [7] Arivazhagan, S., Mepco Schlenk Eng. Coll., Sivakasi, India Priyadharshini, R.A., Seedhanadevi, S. Object, Recognition based on gabor wavelet features, IEEE, March, 2012
- [8] Python 基础教程 Magnet Lie Hetland 人民邮电出版社
- [9] fann 库参考 leenissen.dk/fann/html
- [10] OpenCV 库 2.4.9
- [11] Fann 2.2.0, 2.1.0
- [12] python 2.7