

多媒体信息处理课程

实验报告

霍超凡

1. 实验概述

本次实验聚焦于多媒体信息处理中的多媒体数据检索部分，实现了两个检索小程序，分别用于对音频的检索和图片的检索，伴随着课程的进度，曾经多次更改实验题目，最终受到课程结课考试中关于基于多媒体内容检索的那道题的启发，从而打算实现题目中对音频的检索。一个完整检索系统的搭建过程是非常复杂的，本次实验仅仅聚焦于其中的检索核心方法，而对于系统界面、系统效率等方面没有做出要求。准备实验的期间准备了很多材料，课上仅仅讲到方法，在实践层面介绍的很少，以至于实验无从下手，好在外国有优秀的精品课程可供参考，我找到 Pelloponese University 的《Multimodal Information Processing and Analysis》课程的 PPT 和相关的编程 tutorials[3]，这个课程具有相当大的参考价值，内容丰富，pyAudioAnalysis[4]这个 python 库就是这门课的教授 Theodoros Giannakopoulos 编写的，曾经查过 Theodoros Giannakopoulos 的相关资料，他主要的研究方向在于音频处理方面，编写过一些 Matlab 和 python 的音频处理程序包。实验有关特征提取部分的程序全部调用的是他写的库程序。

报告按照以下内容组织，在搜集相关资料的同时，接触到除课程之外的许多信息，这些信息大都集中于音频信息处理方面，考虑到我们人工智能方向的培养方案中没有关于音频处理方面系统的课程，借此机会补充相关知识。在学习《Multimodal Information Processing and Analysis》编程资料时，我把示例程序手动跑了一遍，学习到如何从音频中提取特征和如何将特征应用于不同的任务中。原先计划做一些关于音频分类的相关工作，可惜时间不够了，只是了解到音频分类的一些方法，这一部分将在第二节介绍。在学习到如何从数据中提取特征之后，我开始搭建自己的检索系统，系统的主要部分在于特征的提取，根据特征向量计算两个多媒体数据的相似程度，进而关于相似度排序给出检索列表。关于系统设计与展示方面放在了第三、四节。记得在结课时，老师曾经询问过对课程的感受与建议，当时课上只是随口一说，考试过后，我反思了自己对这门课的认知与体验，在第五节为课程提出了自己的建议。

2. 音频特征与图片特征提取

2.1. 音频特征

音频是一种波，它通过介质传播到人耳，引起了人的感知，从感知的角度看，刻画声音的两个底层特征是声音的大小和尖锐，这对应于声音的能量特征和频率特征，能量反映声音的强弱，声音的强度不能充分反映音频的特征，我们更多的是从频率角度出发，从而繁衍出许多高层特征。将音频许多微小时间段内的频率按照时间顺序排列，我们得到了这段音频的

频谱，在频谱的基础上，引入人耳感知的特性，有 MFCC 特征。在[2]中介绍了很多 pyAudioAnalysis 中使用音频特征，在实验中着重看了频谱提取、MFCC 特征、pitch 提取和节奏 beat 估计。

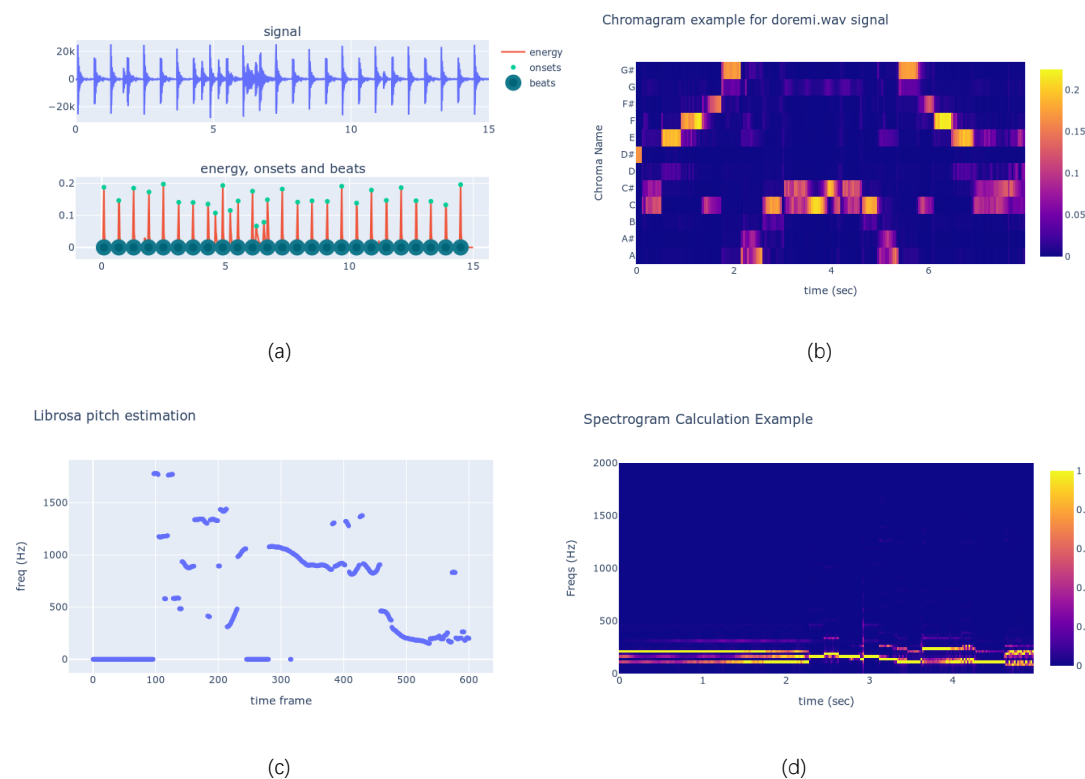


图 1：从音频中提取的不同特征，(a)是节奏估计，(b)是音高，(c)是 pitch 特征，(d)是频谱。

在我们提取完这些特征之后，就可以进行我们下一步识别和分类的工作。[3]中举了许多应用，诸如 Music Speech Classification、Male Female Speech Classification、Music Genre Classification、Audio Event Recognition、Soundscape quality Recognition 等，其中使用的方法大同小异，都是提取特征使用标记数据训练机器学习中的各种分类器。图 2 是其中一个关于音乐风格分类的实例，

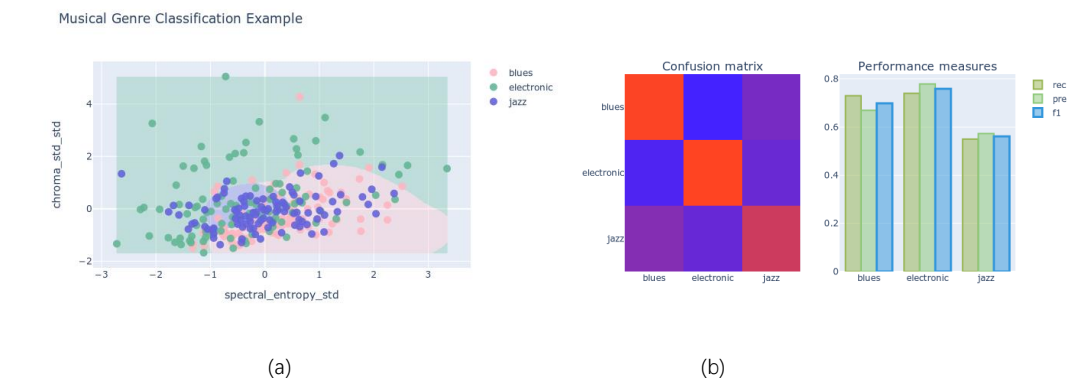


图 2：(a)SVM 分类器分类结果示意图，(b)混淆矩阵和分类精度、召回率。

2.2. 图片特征

图片特征主要使用的是图片颜色特征、LBP 特征、HOG 特征、SIFI 特征，这些特征的提取方法，在《Multimodal Information Processing and Analysis》课程配套的编程 tutorial[3]中均有实现，在实验中特征提取部分，全部照搬了[3]中的代码。

3. 系统设计

系统大致流程如图 3 所示，

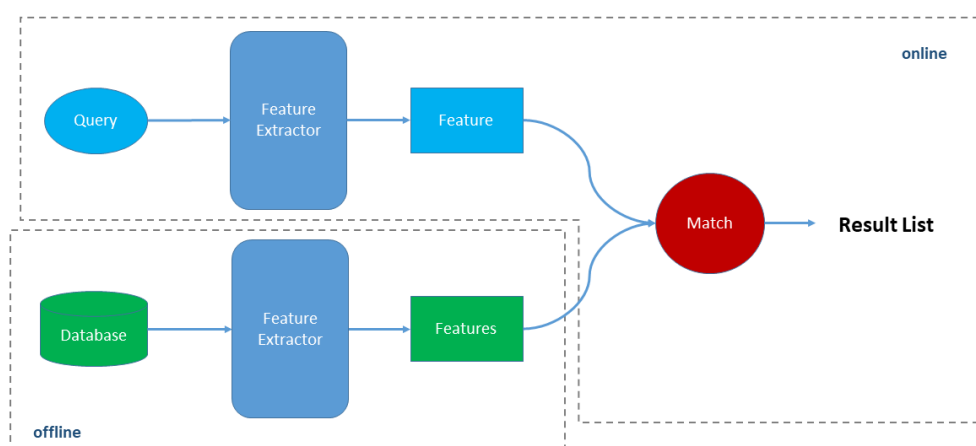


图 3：检索系统流程图

情况下会对数据库中的数据提取特征并建立索引（实验中数据量少、没有建立索引），查询数据可以是图片或者是音频，它结果相同的特征提取器之后得到特征，这个特征和数据库中数据的特征进行比较，计算两两相似度，最后排序输出。实验中本打算使用 django 框架搭建系统，可是这个框架曾经没有使用过，需要现学，后改用 CGI 与 python 程序交互。实验的编程环境在 python 下，使用 html 展示页面、显示结果。

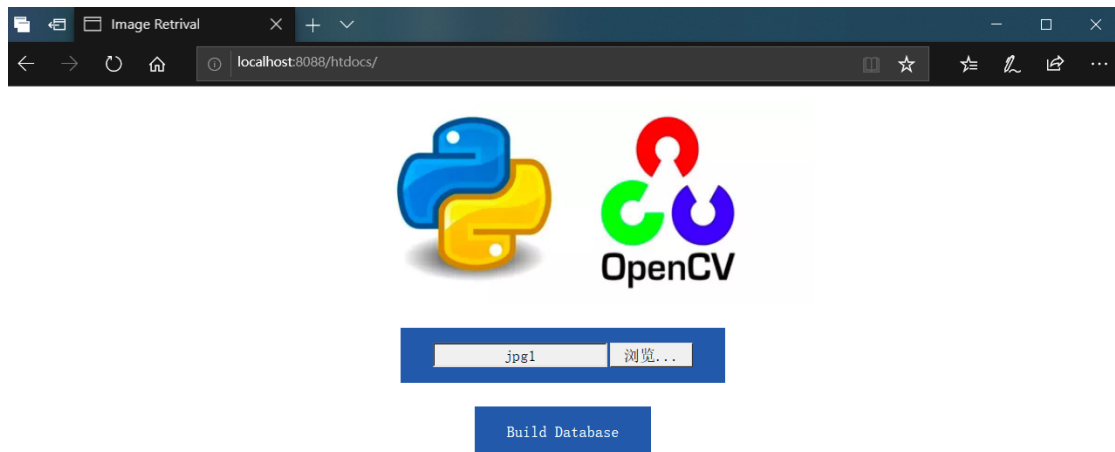
4. 系统界面

如果需要重现系统运行结果，确保本地的环境和我实验中所搭建的环境相同，需要安装 python 相关包（numpy、pandas、pyAudioAnalysis 等）和 CGI，准备数据，数据必须是以多文件夹显示存在，数据库不能只对单个数据文件提取特征。在搭建好环境和准备好数据之后，在实验相关代码文件夹中输入：

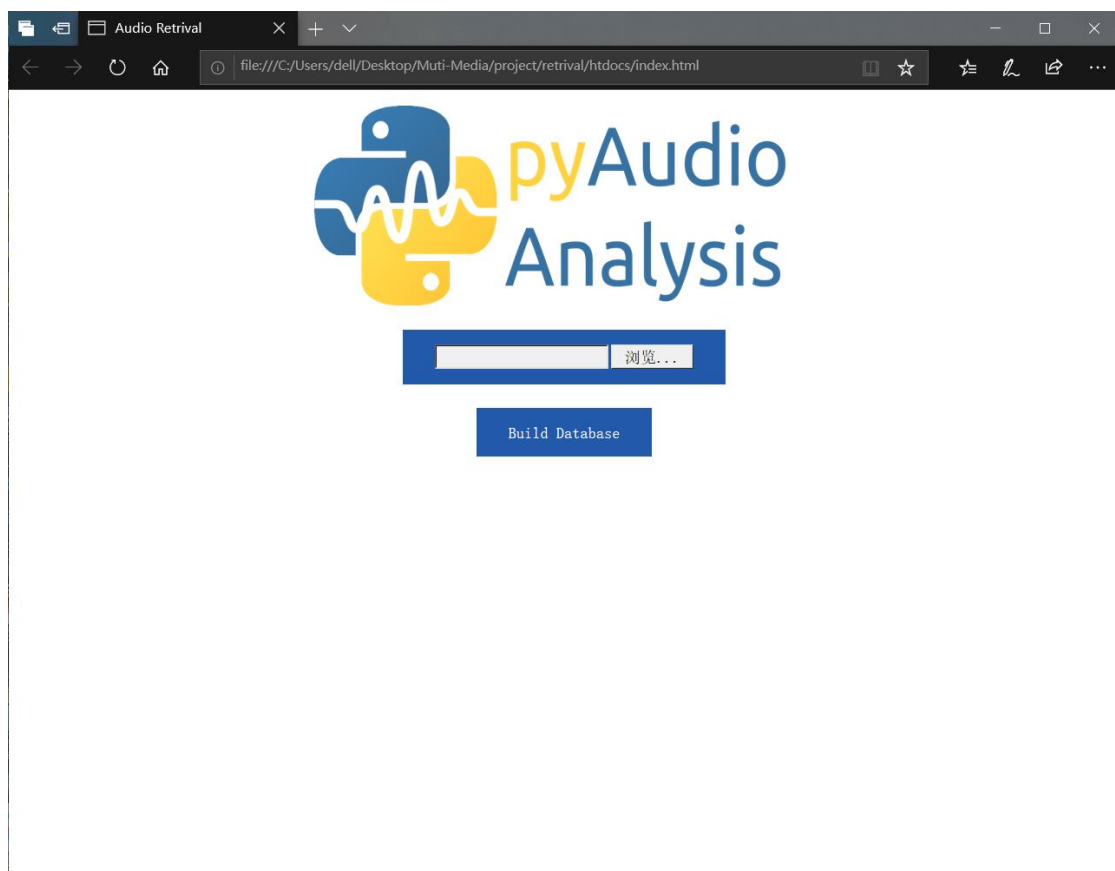
```
python -m http.server --cgi 8088
```

这个命令在本地建立了网页连接，用于 html 前台和 python 后台的交互。在浏览器中打开 <http://localhost:8088>，进入系统，系统界面如下：

对于图片检索，将会看到：

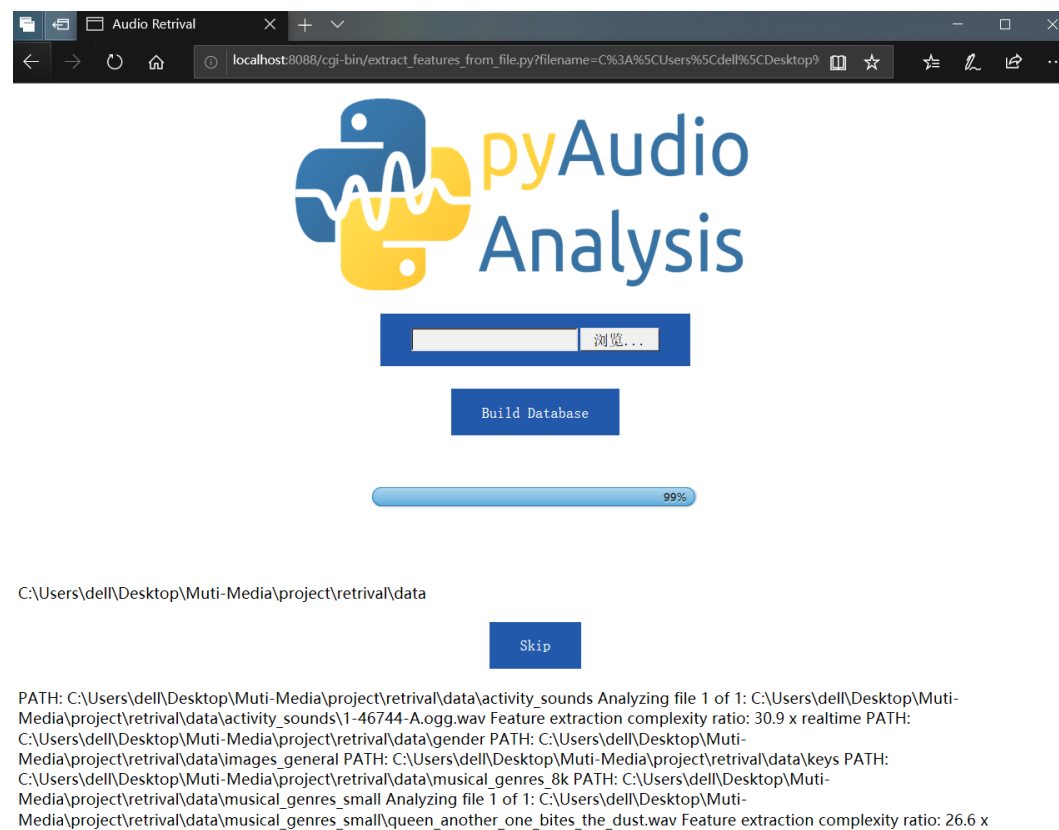


对于音频检索，将会看到：

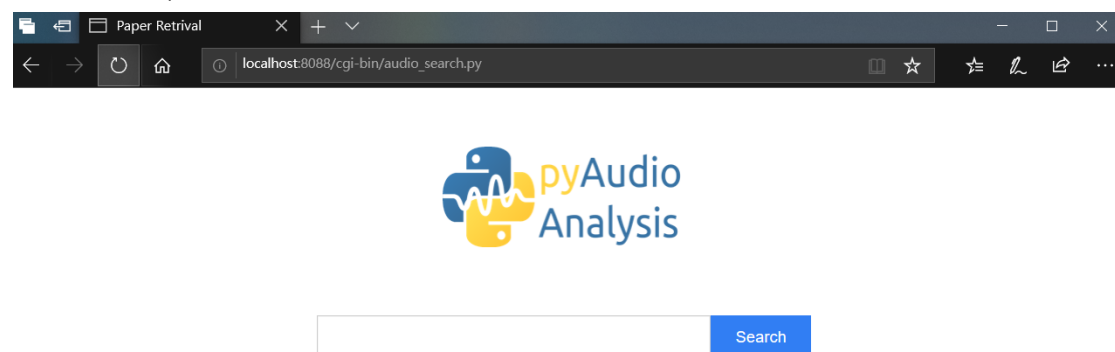


然后，选定事先准备好的数据，建立数据库，提取特征，点击浏览，添加本地路径后，点击

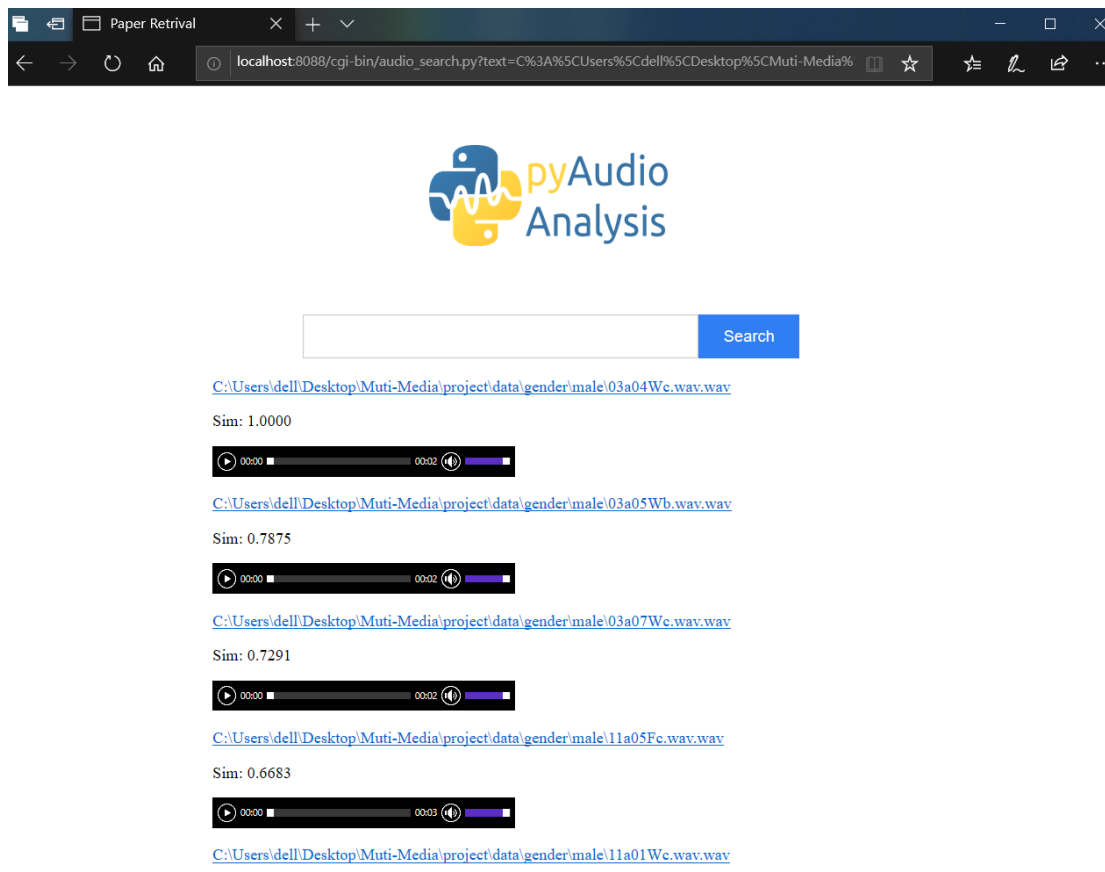
Build Dataset 按钮，当提取完数据库中所有数据的特征后，将会跳转至如下界面：



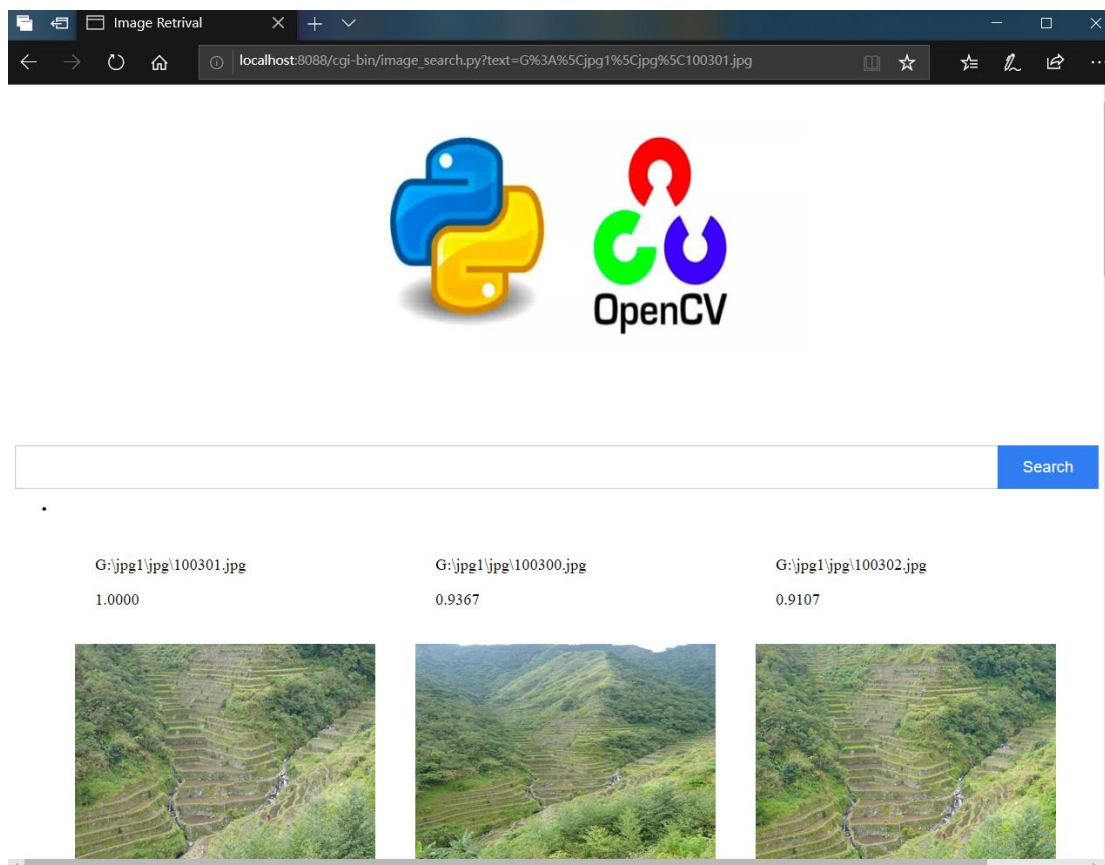
并且在实验程序目录下出现 features.csv，这张二维表存储着从数据库中提取的特征，然后点击 Skip 跳转至检索页面。对于音频检索：



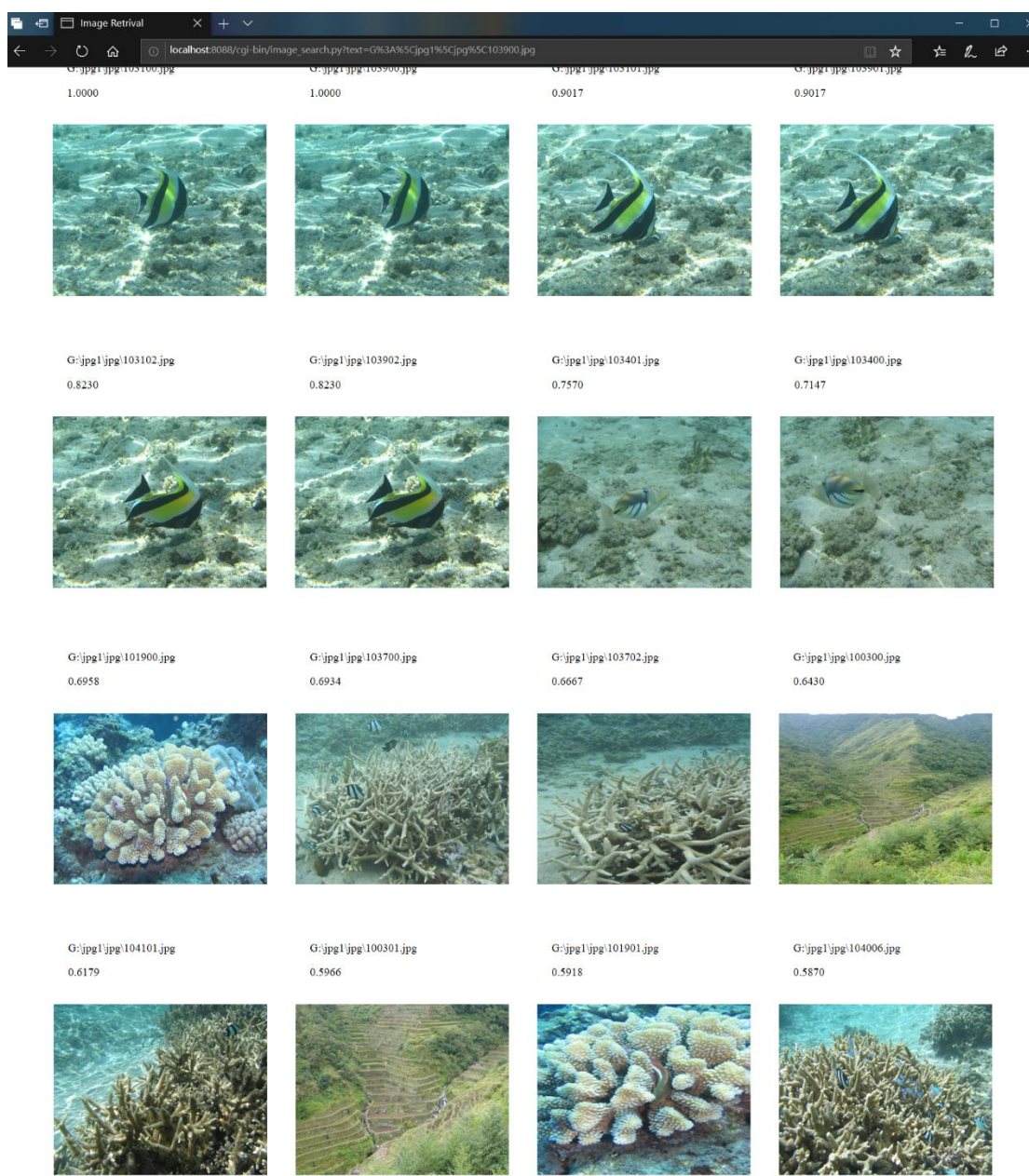
在文本框中输入所需要查询的数据的路径后，点击 Search，对于音频检索：



对于图片检索，



下面给出了其它图片检索的结果,



图片检索所使用的数据库来自于[6], 可以看到检索结果令人满意。

5. 课程体验与建议

多媒体信息处理是一门很硬核的课程, 同时也是人工智能方向同学的一门必修课, 结果这 16 周的学习, 我感觉沉淀在自己脑海里的东西并不多, 这门课不容易讲好, 毕竟达到如 Theodoros Giannakopoulos 水平很难, 课程所涉及的方面也很广, 学起来总感觉学不完。

考虑到我们人工智能方向的培养方案, 在大三上已经学习过关于机器视觉、信息检索的相关内容, 在本学期有开设了 NLP 课程, 在培养方案中关于音频、视频这部分内容有些欠缺, 音频处理中很多基础概念 (如 MFCC) 之前从未接触过, 希望老师在今后的教学过程中加强这部分的教学。关于前面讲到声音的物理特性和光的物理特性的一些细节, 我认为可以淡化, 着重讲声音和光信号特征。

多媒体信息处理同时又是一门实践性很强的学科，我们应该注重些编程方面的教学，实验指导不可或缺，比如介绍 pyAudioAnalysis 类似的库。最后，我们在注重传统方法的同时，也要兼顾目前在深度学习的方法，比如神经网络特征提取器。[3]中附有《Multimodal Information Processing and Analysis》课程的 PPT（Google 网盘链接的形式），可以参考，我认为 Theodoros Giannakopoulos 在内容的组织方面比我们这门课要好，但是这门课缺少我们这门课多媒体信息综合和同步等一些基础性概念。

参考文献

- [1] Choi, Keunwoo & Fazekas, György & Sandler, Mark & Cho, Kyunghyun. (2017). Convolutional recurrent neural networks for music classification. 2392-2396. 10.1109/ICASSP.2017.7952585.
- [2] Giannakopoulos, Theodoros. (2015). pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis. PLOS ONE. 10. e0144610. 10.1371/journal.pone.0144610.
- [3] 《Multimodal Information Processing and Analysis》编程资料，site：
<https://github.com/tyiannak/multimodalAnalysis>
- [4] pyAudioAnalysis, site: <https://github.com/tyiannak/pyAudioAnalysis>
- [5] crnn for audio classification, site: <https://github.com/adityashrm21/music-classification-cnn-pytorch>
- [6] Holidays dataset, site: <http://lear.inrialpes.fr/~jegou/data.php>