**视频生成音乐(视频配乐)技术文档**

夏天 陈长建 方旭东

**一、原理**

我们现在实现的方法大致可分为三步，第一步是关键帧的提取，将视频的识别转化为图像的识别；第二步是通过对得到的关键帧进行情感分析，得到情感坐标；第三步是通过改变基础乐器的组合将得到的情感信息加入，得到最后的音乐。以下是各部分的具体实现。

**视频片段划分：**

通过调用ffmpeg从对每一秒定时提取帧，查看前后两帧的相似性，比较的标准是先将两张图片转换成pixel的list，之后来比较list的编辑距离，计算出两张图片的相似程度。再给定一个判定阈值，当相似度大于该阈值的时候就判断前后两张图片内容发成变化，即视频的节奏发生变化。记录下整个视频中所有内容变化的时间点，精确到秒，这一系列时间点就是视频节奏，同时这些节奏点的帧就是该视频内容的关键帧。

**视频节奏变换特征：**

之前已经采用编辑距离来对视频进行片段划分，现在在每一个片段中，前后帧调用PySceneDetect库再进行一次片段划分，标准为判断其Intensity的绝对值和前后的HSV差值，对视频的前后帧变换再进行一次查找，这样在得到每个片段中的变换次数。最终，该片段的节奏特征由下式计算：

段节奏频率 = 段内变换次数（n） / 段长度（s）

由此得到每一段中的视频节奏特征，用于后面和原先的GoogleNet情感特征进行融合，来得到最终的每一段的段情感特征：

段情感特征 = GoogleNet识别结果 + 段节奏频率 \* 加权系数

加上了段情感特征之后，可以很明显的将视频中的不同乐段的情感特征区分开来，避免之前出现的情感区分度太小的情况。（具体结果参见example）

**图片情感特征：**

情感分析是在已有标定好情感坐标的图片库的基础上，训练一个回归型svm分类器，一次得到新图片的情感坐标。训练svm的方法是先将图片库中的图片通过一个去掉softmax层的GoogleNet网络(相当于对每个图片的预处理)，得到一个3000维的输出，做为图像的特征向量，然后用该3000维特征向量和图片标定的情感坐标对svm进行训练，得到训练好的svm模型。之后对一个新图片，都需要先经过GoogleNet网络之后再用svm分类，得到情感坐标。

情感坐标的生成包括一维和二维。一维的情感坐标是由之前师兄师姐标定好的数据集训练得出的，范围为（-1 ， +1）正表示高兴，负表示不高兴；二维的情感坐标由IAPS数据库中valence-arousal二维的label训练得到，（x,y）的范围也在（-1 ，+1）的正方形平面内。

通过对情感坐标中不同的区域进行划分，将整个情感坐标划分为happy,neutral,sad三种情感，于是通过图片情感分析后的每个关键帧就有了自己的情感。

**音乐伴奏生成：**

之前实现采用两种方法生成音乐，一种是采用RNN对已知的音乐序列建模来生成音乐，另一种是采用根据关键帧的识别情感对基础乐器进行组合来生成音乐。实际实现后发现，第二种对基础乐器组合生成音乐的方法对情感的表达效果更好，并且能够利用上视频节奏变换的信息。

第二种方法的原理：根据已经得到的视频节奏点信息和视频的关键帧情感，提供drum(鼓点),bass(贝斯-低音吉他),high loop(高音循环),acoustic loop(原声循环)，四种不同的组成元素来生成音乐，这四种组成元素相当是音乐中的最小单元，他们由我们自己给出的乐库提供。乐库中每个基础乐段都标定有自己的情感，对于已经得到的视频情感信息和节奏点，给与这四种元素不同的组合，在视频节奏转换的时候用特效来进行过度，生成一个完整的音乐。实际中效果较好的是acoustic loop和drum的组合，其中acoustic loop有全吉他、全钢琴、钢琴吉他混合三种搭配方式，对应生成音乐的GenerateMusic\_all\_guitar.wav,GenerateMusic\_all\_piano.wav, GenerateMusic\_mix\_guitar\_and\_piano.wav，可根据用户喜好自由选择。最后将音乐与原始视频叠加，就得到了含有加上视频节奏变换配乐的视频。

**服务器搭建:**

我们的目的是能搭建一个网页版的API提供给用户使用。由于我们项目所需环境只有在234服务器上有，但是234服务器并没有搭建http服务环境，所以我们采用间接的方式，即先将文件上传到中间服务器上，然后由中间服务器传送给234服务器，以实现上传文件至234服务器上的功能。下面介绍两个服务器的搭建。

1.中间服务器

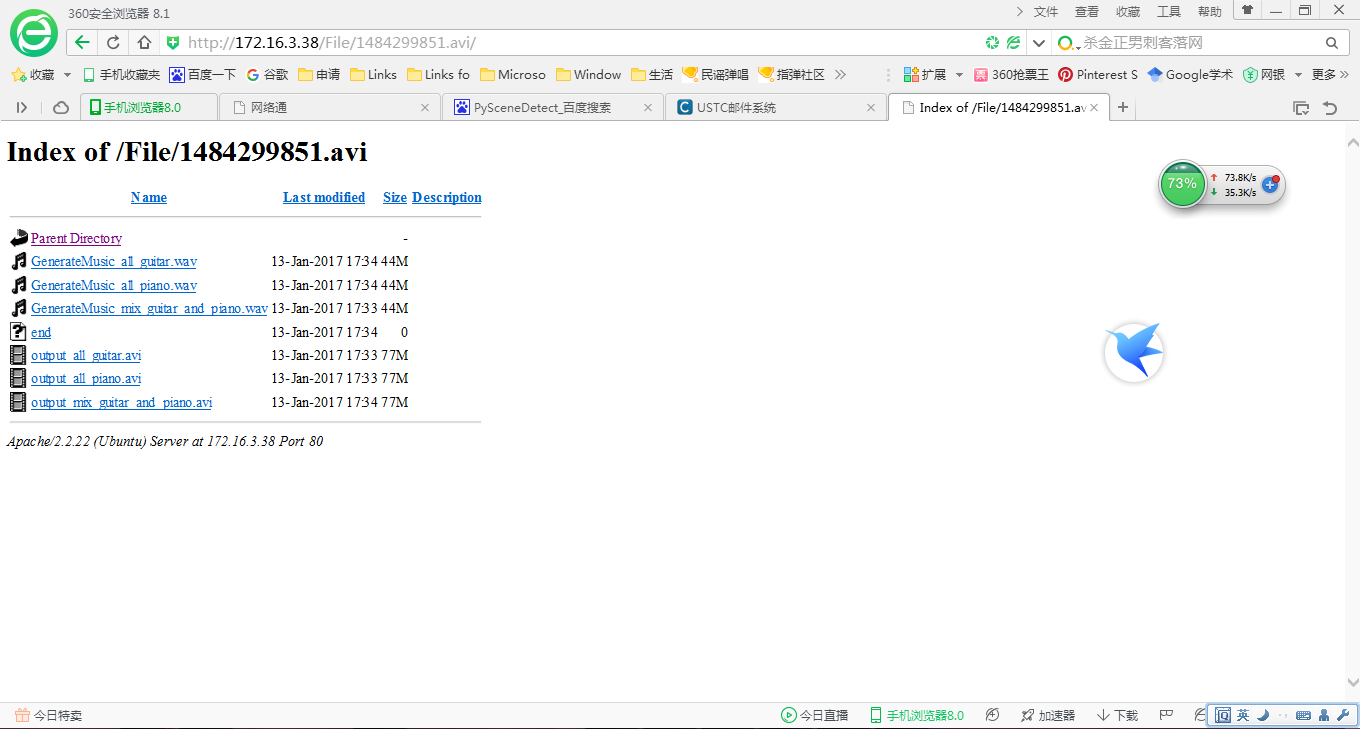
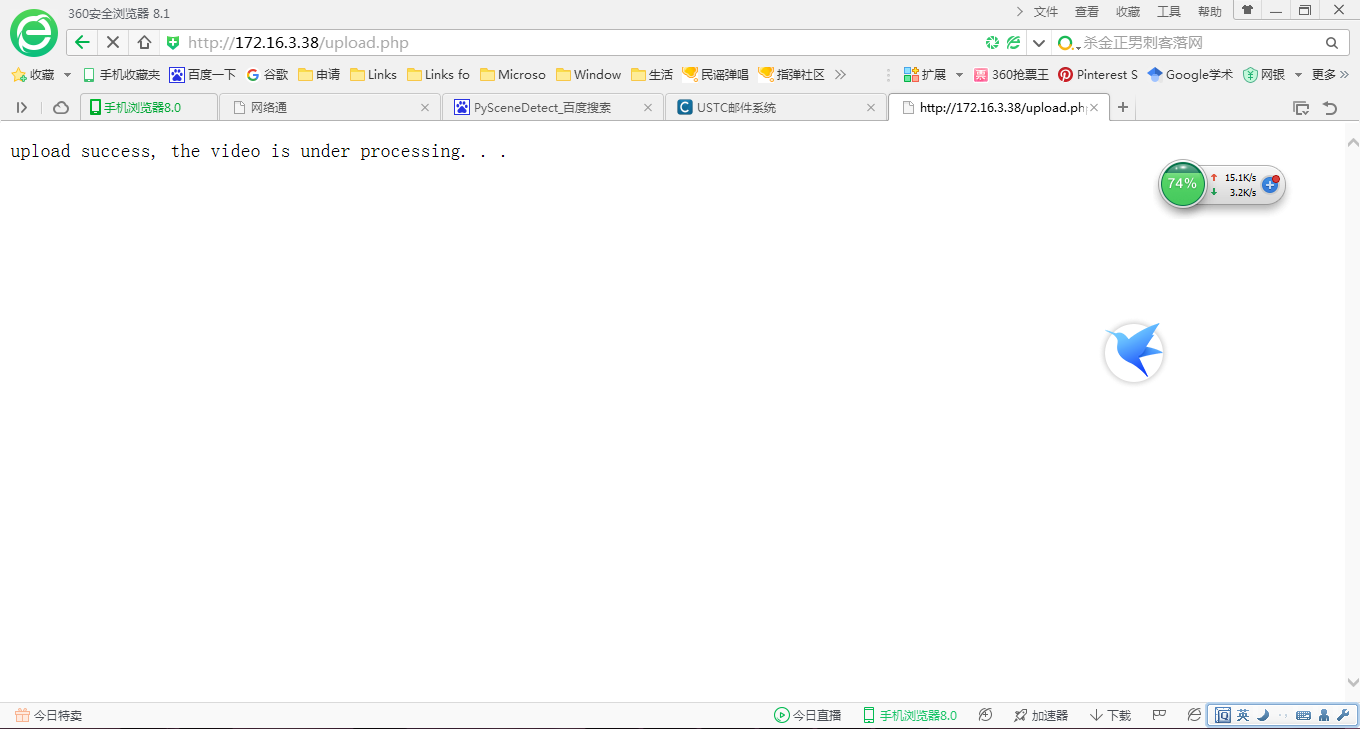
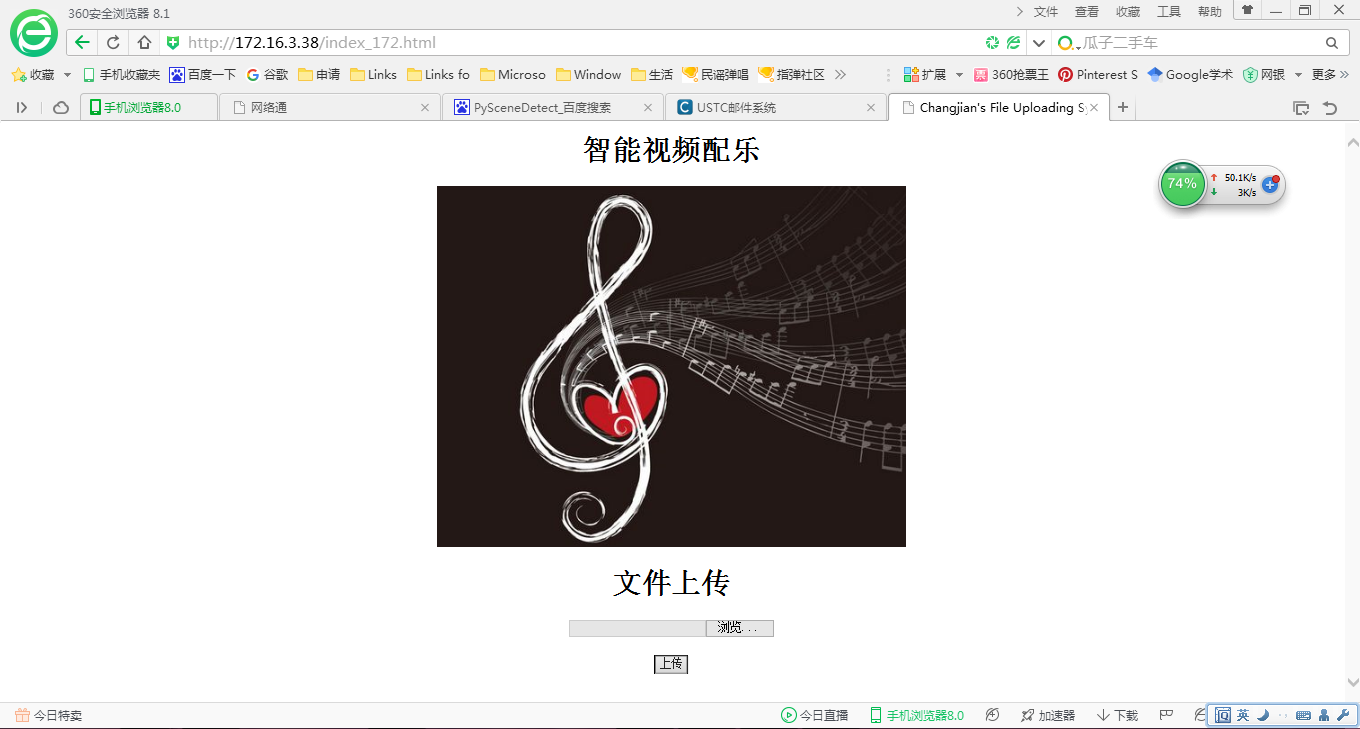
中间服务器我们使用的是学校提供的免费的cloud云主机服务。我们申请的服务器校内IP为172.16.3.38，ubuntu系统，有root权限。我们在服务器上安装apache2来搭建http服务。上传方式为POST，上传的文件默认用时间命名，保存在主目录httpServer文件夹中。

中间服务器后台还运行着一个小脚本，不断检测httpServer内的文件，一旦发现有新的文件存入，则立马将该文件传输至234服务器。传输方式为scp。在使用scp之前可以将两台服务器添加ssh信任，则可以在代码中使用scp而无需身份验证。添加信任的方式就是把对方服务器的ssh公钥添加到自己authorized\_key中。

2.234服务器

234服务器下会以一秒一次的速度检查是否有新文档传入，若有则对传入的文件运行脚本来进行配乐。在完成整个配乐之后再上传到172的服务器上去，然后网页端会进行跳转。

网页端运行的结果如下：



1. **改进及修复**

1.修复了乐段连接的时候偶尔会突然转换而没有过渡音乐分bug

2.解决音频视频长短不一致，导致视频结束音乐一直还在播放的bug

3.略微增加曲库大小，增加happy情感的乐段

4.音频格式由wav转为MP3，减小合成的视频大小，加快传输速度

5.调整鼓声大小随情感强弱而改变

**附：**

**文件列表：**

Python脚本：

compare\_images.py(用于比较视频提取出的帧的不同)

go\_svm.py(对图片进行回归行svm分类，得到一维情感坐标)

go\_svm\_2D.py(对图片进行回归行svm分类，得到二维情感坐标)

googlenet\_custom\_layers.py(googlenet的keras实现)

image\_extract.py(用于调用ffmpeg提取视频帧)

video\_frequence\_detect.py(用于提取视频变换节奏)

music\_function.py(用于调用nsound来生成音乐)

server.py(服务器端接受程序)

权值：

googlenet\_weights.h5(googlenet权值)

svm\_1D\_model(一维svm权值)

svm\_2D\_model(二维svm权值)

示例：

[http://pan.baidu.com/s/1i5mehtb](http://pan.baidu.com/s/1i5mehtb" \t "http://email.ustc.edu.cn/coremail/XT3/mbox/_blank)

**Dependence:**

Nsound (https://github.com/weegreenblobbie/nsound)(开源音乐库，生成音乐时使用)

PySceneDetect(https://github.com/Breakthrough/PySceneDetect/releases)(开源视频处理库)

Sklearn(用于SVM的调用，权值见train\_model.m)

Ffmpeg(视频处理库，用于提取视频的帧)

GoogleNet(权值见googlenet\_weigit.h5)

乐库（由于乐库比较大，放在服务器上的/data/xiatian/src/中）