

# 基于情感语义的短视频配乐APP开发

李倩倩, 陈治灸

贵州大学, 贵州 贵阳  
Email: xxctdys@163.com

收稿日期: 2021年3月22日; 录用日期: 2021年4月15日; 发布日期: 2021年4月22日

---

## 摘要

视频和音乐是互联网上最受欢迎的两种媒体, 两者相辅相成, 并且人类对视频和音乐的感知高度相关。然而给视频内容配上合适的背景音乐, 通常需要一定的专业知识, 普通人大多缺乏这方面的知识, 因此本文通过计算机为音乐和图像搭建桥梁, 开发设计一个给视频匹配背景音乐的手机APP, 旨在给用户视频配上情感相同的音乐, 增加整个视频的合理性。本文主要技术手段是利用视频图像和音乐之间的情感语义相似性, 通过python的kivy框架来设计开发一款跨平台的视频配乐系统。

## 关键词

视频配乐, 软件需求分析, App, kivy

---

# Short Video Soundtrack APP Development Based on Emotional Semantics

Qianqian Li, Zhijiu Chen

Guizhou University, Guiyang Guizhou  
Email: xxctdys@163.com

Received: Mar. 22<sup>nd</sup>, 2021; accepted: Apr. 15<sup>th</sup>, 2021; published: Apr. 22<sup>nd</sup>, 2021

---

## Abstract

Video and music are the two most popular media on the Internet. The two complement each other, and human perceptions of video and music are highly correlated. However, matching the video content with appropriate background music usually requires certain professional knowledge. Most ordinary people lack this knowledge. Therefore, this article builds a bridge between music and images through a computer, and develops and designs a mobile app that matches the back-

ground music to the video. It is designed to match the user's video with music with the same emotion and increase the rationality of the entire video. The main technical means of this paper is to use the emotional semantic similarity between video images and music to design and develop a cross-platform video soundtrack system through the python kivy framework.

## Keywords

Video Soundtrack, Software Requirement Analysis, App, kivy

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

由于数字媒体技术的进步, 数字视频和音乐已经在人类生活中普遍存在。视频被广泛用于记录和捕捉重要的时刻, 如婚礼, 毕业典礼和生日聚会等。此外, 随着 YouTube、Flickr、抖音、快手等社交媒体网站上传的视频内容越来越多, 用户可以轻松自由地分享自己的个人视频, 然而冗长未经编辑的视频看起来很有趣。因此, 让视频更具吸引力逐渐成为人们的基本需求。目前有几种方法可以自动编辑视频使其变得更有趣: 一种是视频摘要, 主要目的是从原始视频中发现重要部分, 自动生成简洁但有信息的摘要帮助用户理解, 并且目前在视频自动摘要方面已经有了大量的研究成果[1]; 另一种方法是在视频中添加背景音乐使得视频更加生动, 基本原理是通过对视觉与听觉跨模态信息的分析显示, 这种视听互补信息在改善多媒体语义信息理解方面具有重要的利用价值[2]。视频和音乐都是被广泛使用的媒体, 但它们之间的联系并没有被很好地探索。有效的视频与音乐匹配技术在跨模态检索、音乐检索以及自动生成音乐视频等方面都有广泛的应用。

视觉刺激会使人们在听音乐时产生共鸣, 而音乐则会在视觉感知中为场景增色。仅有音乐或只有视频图像给人的感觉较为乏味, 但当图像或剪辑的视频与音乐同时出现时将带来人们更好的视听享受。因此对于现代电影和电视节目, 音乐是重要组成部分。而如何将音乐和视频图像更好的匹配融合, 成为时代发展的研究课题。许多研究人员从不同角度开展了诸多的研究, 比如 Olivier collignon 等人用自我报告和对比反应时间的方法证明视听情绪内容一致的时能提升带来的情绪的感受, 当不一致时视听内容的表达将会产生相互抑制的影响[3]。

用户在与他人分享精彩视频之前, 需要进行视频编辑和气氛渲染, 选择一种合适的背景音乐嵌入到所拍摄视频中成为最有效的方式之一[4]。因此, 人们强烈需要一种自动背景音乐推荐 APP, 从而在显著减少用户工作量的同时, 也使得用户可以选择和视频情感更加匹配的音乐而不是盲目跟风选择当下流行音乐。合适的背景音乐能更好渲染视频想要表达的情感, 且目前短视频行业经常出现某些音乐滥用的问题, 一首音乐大火之后, 会出现大量用户选择跟风使用此音乐, 而不在于此音乐与自己视频内容是否相匹配, 使得同质化视频越来越多。总之, 用户在制作短视频时, 多局限于使用出现次数较多的背景音乐, 缺乏新意。长此以往, 同质化的短视频在平台上不断传播, 造成用户审美疲劳甚至产生厌烦, 其观看短视频满意度会大幅度。针对以上问题, 本文根据视频图像和音乐之间的情感语义相似性, 通过 python 的 kivy 框架来设计开发一款跨平台的视频配乐软件, 主要目的是使用户在为视频选择背景音乐时有更加多元化的选择, 帮助用户打造不落俗套的作品。

## 2. 软件需求分析

在软件设计阶段和软件生命周期中, 需求分析是非常重要的阶段。本文设计并开发了一种识别并推荐相应情感音乐的配乐软件, 该软件能够对图片或视频中的情感进行识别, 并将识别到的情感类型进行分类, 用户可选择心仪的情感类型, 系统依据用户的选择推荐相应情感类型的音乐供给用户选择, 从而达到为视频或图片匹配符合情感配乐的目的。软件设计涉及到了视频解码、情感识别、历史记录展示、音乐推荐等功能。需求分析的内容是为要开发的软件提供完整, 清晰和特定的要求, 并确定软件需要哪些功能来实现相应任务。在需求调研阶段, 研究者深入挖掘了用户场景, 发现用户的确定需求和潜在需求从而确定软件各功能, 并根据四象限分析法确定各需求优先级, 其中 P0 优先级最高。具体需求及功能如表 1。

**Table 1.** Specific needs and functions

**表 1.** 具体需求及功能

模块	场景	需求	功能	优先级
首页	用户一进入 APP 可以看到视频配乐的操作区域	配乐需求	导入视频, 情感识别, 情感选择, 配乐推荐, 配乐选择	P0
	用户一进入 APP 可以看到拍摄视频的操作区域	现场拍摄视频需求	现场拍摄视频	P0
视频剪辑	用户希望看到以往的操作记录	历史记录展示需求	展示历史记录	P0
	用户希望对配乐完成的视频进行剪辑美化	视频剪辑美化需求	添加滤镜, 文本编辑, 拆分, 贴纸, 音量大小	P1
视频发布	用户希望将配乐剪辑完成的视频发布在平台上	发布需求	在平台上发布视频	P1
消息	用户希望看到其他用户和自己互动的信息	互动需求	评论回复我的, @我的, 赞, 粉丝, 系统消息	P1
	用户想修改个人资料	互动需求	编辑个人资料	P1
我的	添加好友, 让别人了解我	互动需求	添加好友	P2
	查看曾经发布过的视频	作品数据统计需求	发布过的作品, 作品的观看次数	P2

### 2.1. 软件框架构建

本文软件框架如图 1 所示, 整个软件系统主要包含两个主要资料库: 视频图像数据库和音乐资料库。在构建视频图像数据库时, 使用差分法[5]提取视频关键帧来代表整个视频。并且邀请多名相关专业背景的师生为两个数据集进行了标注, 取得视频图像和音乐的情绪标签, 且为每个视频提取的关键帧图像配五首歌曲, 使得每一组匹配的视频图像 - 音乐对使用一类共同的情感标签。随后用神经网络对成对的视频图像 - 音乐数据进行训练, 建立起视频图像 - 音乐的关联模型, 实现为视频图像推荐情感相似的音乐。具体来说, 将提取的关键帧序列图片进行尺寸归一化处理, 然后使用两个独立 VGG19 网络分别提取视频图像和音乐的深度特征, 并将视频图像和音乐的特征映射到共同的语义空间, 采用相似性度量发掘视频图像和音乐之间的关联, 进而进行视频图像和音乐之间的互检。用户模块是用户使用本软件的流程, 用户输入自己的视频, 软件会先提取视频的关键帧, 然后提取关键帧的特征, 接着识别关键帧的情感, 输出推荐的背景音乐列表, 用户可以在推荐的列表内挑选出自己认为更匹配的音乐。

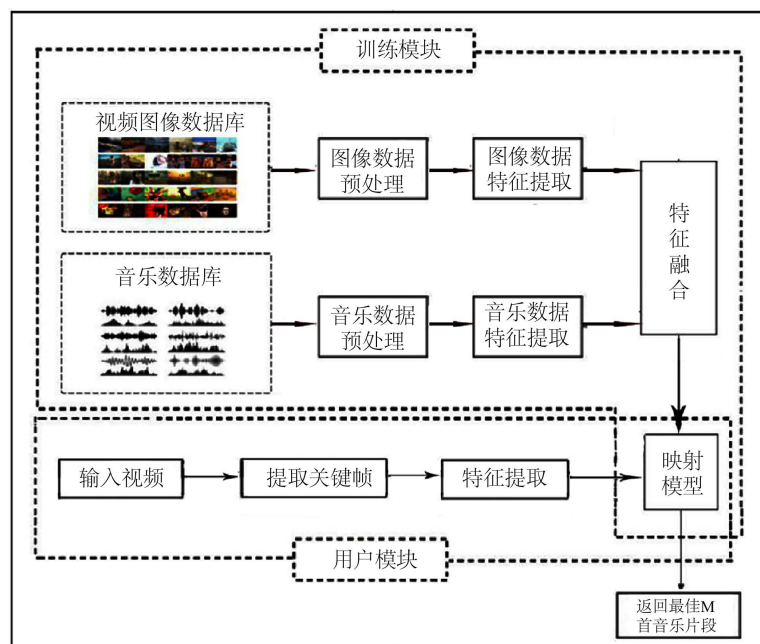


Figure 1. Software framework diagram  
图 1. 软件框架图

## 2.2. APP 核心功能业务流程

根据软件自身定位可知, 给视频配乐是此软件的核心功能, 因此在软件开发阶段应首先考虑此功能的开发设计, 给视频配乐的业务流程如图 2 所示。

## 3. APP 技术实现

### 3.1. 开发框架

本文的 APP 在 Linux 系统中完成开发, 基于 python 的 Kivy 框架设计。Kivy 是基于 python 的一个开源的 GUI 图像用户界面(Graphical User Interface, 简称 GUI)工具, 用来快速创建跨平台的 python 应用[6]。由于其跨平台性, kivy 开发的软件可以在 Windows, Linux, Mac, Android 和 IOS 设备上运行。Kivy 框架是用 Python 和 Cython 编写, 使用 Cython 来加速与图形渲染, 矩阵计算以及事件分派和属性操作有关的所有组件。硬件加速则是通过 OpenGL ES 2 实现。Kivy 提供了一种 kv 语言, 使得界面 UI 文件和程序文件可以相互分离, 界面 UI 文件在扩展名为.kv 的文件中创建, 用户可以独立于 python 代码修改调试应用程序的 GUI 界面, 并且应用程序的代码也更清晰易懂。相比较于其他图像界面框架 Tkinter, PyQt, PyGTK 等, kivy 兼容性强, API 文档和开发教程丰富。因此本文选择用 Kivy 实现基于情感语义匹配的视频配乐软件。

### 3.2. 打包工具

本文在 python-for-android(p4a)编译, p4a 是一个开源构建工具, 可将 python 代码打包到独立的 android APK 中, 使用 p4a 需要手动下载 Android SDK 和 Android NDK 的安装包, 随后安装构建工具, 配置 SDK 和 NDK 的环境变量。Kivy 的 APK 就是常规的 Android APP, 和其他的 APP 一样可以四处发布, 比如谷歌 Play 商店等等。p4a 工具最初是为了 kivy 跨平台图像框架开发的, 但现在支持多种前端框架, 也可以轻松扩展打包用到其他类型的 Python 应用程序 Android app 上。

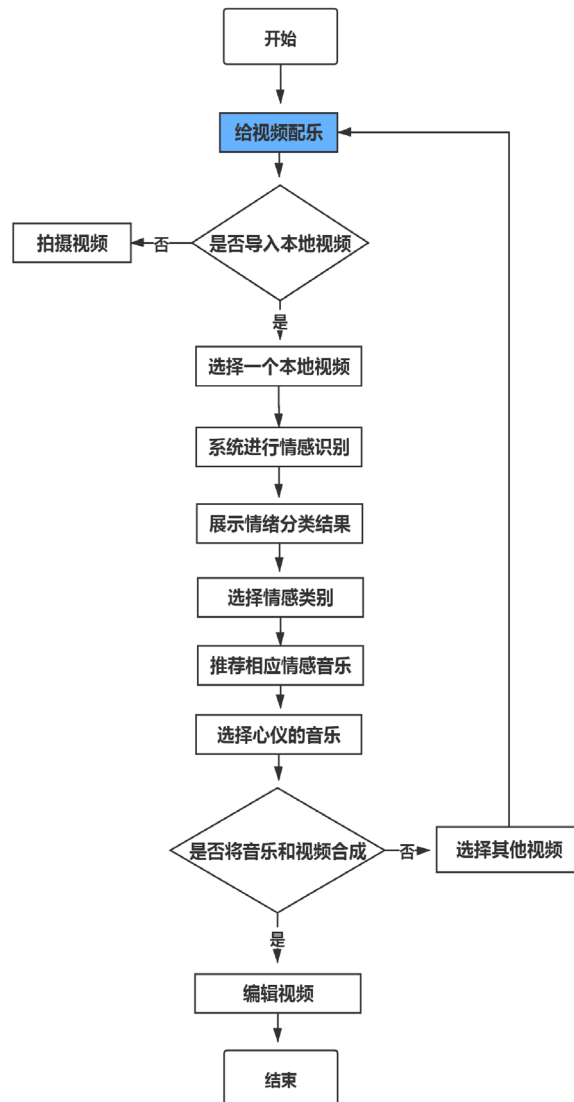


Figure 2. Core function business process  
图 2. 核心功能业务流程

### 3.3. APP 项目打包流程

将程序打包为 apk, 首先要先创建项目文件夹, 文件夹里主要包含程序文件 main.py 和 UI 界面文件 MYAPP.kv。之后创建 p4a 打包的配置文件 “.p4a”, 生成的.p4a 配置文件, 代码如下:

```
———dist_name ccw  
———android_api 19  
———sdk_dir/home/kivydev/andr/android-sdk-linux  
———ndk_dir/home/kivydev/andr/crystax-ndk-10.3.2  
———requirements python3, kivy tensorflow, numpy, keras  
———private.  
———package com.myapp.ccw  
———name ccw
```

——version 1.0

——bootstrap sdl2

随后通过命令“p4a apk”打包 apk，完成打包后会生成 apk 文件，把 apk 文件复制到共享文件夹，在手机或安卓模拟器中安装即可。

## 4. APP 展示

本节展示 APP 成品，主要包括用户使用流程和页面展示。关于使用流程，用户可自行选择上传本地视频或者拍摄视频，随后系统对上传视频进行情感识别，得到视频情感后，推荐情感相似音乐，用户可根据系统推荐的音乐选择心仪合适的。由于时间和技术等原因，我们目前只实现了一部分功能，实现的部分功能如图 3 所示，主界面包括上传本地视频功能键，拍摄视频功能键以及历史记录模块；过渡页面是系统对上传的视频进行情感识别，页面中会出现识别进程以提示用户；识别出视频情感后，进入识别情感类别页，在此页面用户可以在识别的情感类别里选择自己认为更符合视频情感的情感类别，之后进入推荐歌曲页，在此页面会显示 APP 根据用户选择的情感类别所推荐的音乐，用户可根据其推荐音乐自行选择。并且本 APP 也可以查询历史记录，进入历史查询页面可以查询自己以往的视频配乐记录，便于用户操作。

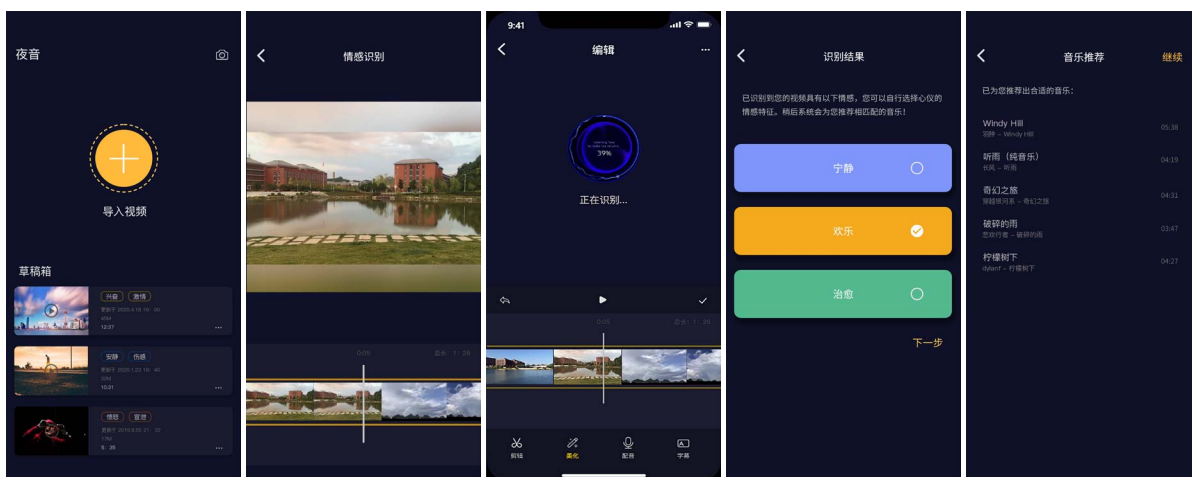


Figure 3. APP interface

图 3. APP 界面

## 5. 总结

本文基于 python 语言，利用 kivy 跨平台图形框架为短视频配乐开发了一款基于 Android 移动终端的配乐 App，得益于 kivy 对快速开发的支持，本 App 可以快速实现。本文软件在 Linux 系统中完成开发，然后移植到 Android 手机上，大大提高了软件的实用性。基于情感语义的短视频配乐 APP 优点如下：一方面可实现视频自动配乐，减少用户编辑短视频的工作量，并且基于情感进行短视频配乐能更好的渲染视频得到用户想要表达的效果；另一方面为用户提供了一种有效手段去改善短视频背景音乐单一，缺乏新意的行业现状。

## 参考文献

[1] 武光利, 李雷霆, 郭振洲, 王成祥. 基于改进的双向长短期记忆网络的视频摘要生成模型[J]. 计算机应用, 2021:

- 1-9.
- [2] 张华, 蒋玉暎, 蒋伟, 张晶晶. 基于情感的视听关联问题[J]. 中国传媒大学学报(自然科学版), 2017, 24(1): 35-40+50.
  - [3] Jeong, J.W., Diwadkar, V.A. and Chugani, C.D. (2011) Congruence of Happy and Sad Emotion in Music and Faces Modifies Cortical Audiovisual Activation. *Neuroimage*, **54**, 2973-2982.  
<https://doi.org/10.1016/j.neuroimage.2010.11.017>
  - [4] 徐鸿雁. 基于视频场景分析的背景音乐自动推荐方法[J]. 计算机应用, 2014, 34(S1): 268-269+326.
  - [5] 王智文, 王宇航, 蒋联源, 张灿龙, 欧阳浩, 黄镇谨. 基于关联帧差分法的运动目标检测与跟踪[J]. 现代电子技术, 2021, 44(2): 174-178.
  - [6] 王义锋. 面向家庭环境的人体行为分析方法研究及软件开发[D]: [硕士学位论文]. 杭州: 浙江工业大学, 2020.