

数字视音频卷子试做

数字视音频卷子试做

16-17

第一部分选择题，7个题10个空，10分

- 1.音乐制作的步骤
- 2.midomi
- 3.语音发声形成过程四个空。具体什么部位影响声音的什么特征
- 4.特征提取预处理的目的
- 5.关键帧提取算法
- 6.镜头相似度提取的特征
- 7.视频目录构造的过程

第二部分判断题，3个题，每题4分

第三部分，简答题，6道题，每题6分

- 1.答旋律节奏和声以及这三个的定义
- 2.三种音乐编程方法
- 3.说话人识别过程中的假设和简化
- 4.语图生成的过程
- 5.运动向量搜索的三种方法
- 6.时序结构图的构造

第四部分，论述题，三个题，共42分

17-18

选择题

- 1.音乐编程（API）
- 2.MIDI 选个错误的 选项看起来有点陌生
- 3.说话人识别随时间的演变 3空
- 4.预处理中加重高频段
- 5.忘记了 有没有人来补充的
- 6.哪个不是关键帧提取算法
- 7.视频目录构造过程 排序

判断题

- 1.有板有眼 旋律
- 2.MFCC只能获取说话人的身份信息
- 3.视频压缩和播放顺序一定相同

简答题

- 1.音乐的软件编程 合成顺序
- 2.音乐三种高级表现形式
- 3.语音时域频域分析的特点和变量
- 4.给出机器能够与人对话图 其中所用到的知识
- 5.时序结构图场景检测
- 6.IPB帧

论述题

- 1.midomi和sha什么公司 哼唱和乐纹检索区别和技术啥的
- 2.声纹识别和乐纹识别的联系
- 3.渐变镜头的数学模型推导

16-17

第一部分选择题，7个题10个空，10分

1.音乐制作的步骤

作词作曲 编曲 录音 混音

2.midomi

基于轮廓因子索引的快速检索法

3.语音发声形成过程四个空。具体什么部位影响声音的什么特征

声门振动的快慢，决定声音的基本频率（即音高）。

口腔、鼻腔、舌头的位置、嘴型等，决定声音的内容（即音色）。

肺部压缩空气的力量大小，决定音量大小

4.特征提取预处理的目的

特征提取中的预加重，保留一定频率范围的信号，实际上是一个高通滤波，对高频信号有着很好的放大作用，一定程度上可以消除唇齿效应。

5.关键帧提取算法

基于聚类的关键帧提取

镜头边界法

特征转变法

运动分析法

6.镜头相似度提取的特征

空间特征（颜色相似度等）

时间特征（运动相似度）

7.视频目录构造的过程

- 镜头边缘检测
- 关键帧提取
- 时空特征提取
- 时间可适性成组
- 场景结构构造

第二部分判断题，3个题，每题4分

1.MIDI

2.MFCC

3.Group的定义

第三部分，简答题，6道题，每题6分

1.答旋律节奏和声以及这三个的定义

节奏：组织起来的音的长短关系

旋律：长短，高低，强弱不同的一连串乐音有组织的进行

和声：和弦+和声进行

2.三种音乐编程方法

Nyquist、OpenAL、Matlab、Flash Action Script

3.说话人识别过程中的假设和简化

短时平稳假设

使得数学模型简化：

1. 当前状态只与前一状态有关，而与更早的状态无关
2. 当前状态的输出只与当前状态有关，而与其他任何状态均无关

4.语图生成的过程

将声音分成很多帧，每帧对应一个频谱，将其映射到灰度表示，添加时间维度，得到随时间维度变化的频谱图。

5.运动向量搜索的三种方法

顺序搜索

2D对数搜索

层次搜索(Hierarchical Search)

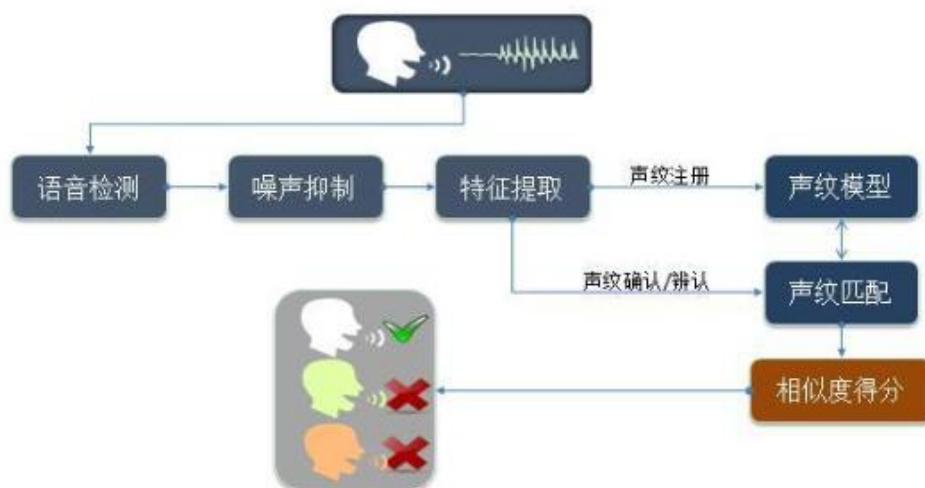
6.时序结构图的构造

- 视频解码
- 视频切分
- 关键帧提取
- 视频聚类分析
- 构造时序图
- 按照时序图浏览

第四部分，论述题，三个题，共42分

1.实验的五个题目选一个说明算法流程和关键技术

2.说话人识别流程，以及结合各种图表（元音三角形、语图...）说明得分低的音频的特点



3.I、P、B定义，P帧视频压缩方法

I帧(Intra-Frame)是帧内压缩，不使用运动补偿，提供中等的压缩比。由于I帧不依赖于其他帧，所以是随机存取的入点，同时是解码中的基准帧。

P帧(Predicted-Frame)根据前面的I帧或P帧进行预测，使用运动补偿算法进行压缩，因而压缩比要比I帧高，数据量平均达到I帧的 1/3左右。P帧是对前后的B帧和后继的P帧进行解码的基准帧。P帧本身是有误差的，如果P帧的前一个基准帧也是P帧，就会造成误差传播。

B帧(Bidirectional-Frame)是基于内插重建的帧，它基于前后的两个I、P帧或P、P帧，它使用双向预测，数据量平均可以达到I帧的 1/9左右。B帧本身不作为基准，因此可以在提供更高的压缩比的情况下不传播误差。需要指出的是，尽管我们使用帧 (Frame) 这个词，但是MPEG2本身没有规定进行数字图像压缩时必须使用帧作为单位，对于隔行的视频图像，可以使用场 (Field) 作为单位。

17-18

选择题

1.音乐编程 (API)

OpenAL

2.MIDI 选个错误的 选项看起来有点陌生

3.说话人识别随时间的演变 3空

4.预处理中加重高频段

5.忘记了 有没有人来补充的

6.哪个不是关键帧提取算法

7.视频目录构造过程 排序

判断题

1.有板有眼 旋律

2.MFCC只能获取说话人的身份信息

3.视频压缩和播放顺序一定相同

简答题

1.音乐的软件编程 合成顺序

overture SONAR Adobe audition

Nyquist OpenAL Matlab

作词作曲 编曲 录音 混音

2.音乐三种高级表现形式

节奏 旋律 和声

3.语音时域频域分析的特点和变量

时域分析：简单直观，清晰易懂，物理意义明确；音量、过零率、音高 (pitch)

频域分析：具有实际物理意义；共振峰参数，音高/基音周期(pitch) MFCC LPCC

4.给出机器能够与人对话图 其中所用到的知识

- 语音学与音系学—研究语言的语音
- 形态学—研究词的有意义的组合
- 句法学—研究词与词之间的结构关系
- 语义学—研究意义
- 语用学—研究如何用语言来达成一定的目的
- 话语学—研究大于段的语言单位

5.时序结构图场景检测

视频解码

视频切分

关键帧提取

视频聚类分析

构造时序图

按照时序图浏览

6.IPB帧

I帧(Intra-Frame)是帧内压缩，不使用运动补偿，提供中等的压缩比。由于I帧不依赖于其他帧，所以是随机存取的入点，同时是解码中的基准帧。

P帧(Predicted-Frame)根据前面的I帧或P帧进行预测，使用运动补偿算法进行压缩，因而压缩比要比I帧高，数据量平均达到I帧的 1/3左右。P帧是对前后的B帧和后继的P帧进行解码的基准帧。P帧本身是有误差的，如果P帧的前一个基准帧也是P帧，就会造成误差传播。

B帧(Bidirectional-Frame)是基于内插重建的帧，它基于前后的两个I、P帧或P、P帧，它使用双向预测，数据量平均可以达到I帧的 1/9左右。B帧本身不作为基准，因此可以在提供更高的压缩比的情况下不传播误差。需要指出的是，尽管我们使用帧(Frame)这个词，但是MPEG2本身没有规定进行数字图像压缩时必须使用帧作为单位，对于隔行的视频图像，可以使用场(Field)作为单位。

论述题

1.midomi和sha什么公司 哼唱和乐纹检索区别和技术啥的

2.声纹识别和乐纹识别的联系

3.渐变镜头的数学模型推导

Fade /Dissolve 镜头产生模型

所谓“fade”，就是原始视频帧像素点增强或减弱的一个过程。假设原始视频帧序列为 $S_1(x, y, t)$ ($t \in [0, T]$)，那么视频序列 $S(x, y, t)$ 称为“fade”序列，是指使用一个时间单调函数 $f(t)$ 与 $S_1(x, y, t)$ 相乘： $S(x, y, t) = f(t) \cdot S_1(x, y, t)$ 。如果 $f(0) = 0$ 和 $f(t) = 1$ ，那么就叫做“Fade in（淡入）”；如果 $f(0) = 1$ 和 $f(t) = 0$ ，则就叫“Fade out（淡出）”。

所谓“Dissolve”，就是视频场景A逐渐变成没有后，然后视频场景B逐渐显示出来的过程。可以如下表示Dissolve中视频图像序列 $S_n(x, y)$ ：当 $0 \leq n \leq L_1$ 时， $S_n(x, y) = f_n(x, y)$ ；

当 $L_1 \leq n \leq L_1 + F$ 时， $S_n(x, y) = (1 - \frac{n - L_1}{F})f_n(x, y) + \frac{n - L_1}{F}g_n(x, y)$ ；

当 $L_1 + F \leq n \leq L_2$ 时， $S_n(x, y) = g_n(x, y)$ 。其中， $f_n(x, y)$ 表示视频场景A、 $g_n(x, y)$ 表示视频场景B、 L_1 表示场景A持续时间、 L_2 表示场景B持续时间、 F 表示场景A和场景B或多或少重叠在一起的Dissolve持续时间。

基于如上的产生模型，可以推导出 Fade/Dissolve 等镜头的固有特性。假设镜头场景 $f_n(x, y)$ 和镜头场景 $g_n(x, y)$ 的均值与方差分别是 m_f 、 m_g 、 σ_f^2 和 σ_g^2 ，则 $S_n(x, y)$ 的均值 $m_{s,n}$ 为：当 $0 \leq n \leq L_1$ 时， $m_{s,n} = m_f$ ；当 $L_1 \leq n \leq L_1 + F$ 时，

$m_{s,n} = [m_f - \frac{L_1}{F}(m_g - m_f)] - \frac{n}{F}(m_f - m_g)$ ；当 $L_1 + F \leq n \leq L_2$ 时， $m_{s,n} = m_g$ 。从这里

可以看出，在 Dissolve（或 Fade）中，视频帧灰度值的均值是一个线性函数，灰度均值线性增加（或减少）到某个值。更进一步，视频帧灰度值均值的一阶导数是常数。P