PAJRD20170275---一种融合音频情感和视频表情分析的实时欺诈识别系统

背景技术：

在反欺诈识别中，对语音中隐含的情绪进行分析难度较大，目前还没有成熟稳定的技术应用于语音分析以识别欺诈,。

微表情识别比较完善，但是实际生产样本数据很难收集，提取欺诈识别相关的微表情特征比较困难，实际效果堪忧。

目前市场上还没有同时进行语音和表情分析的欺诈识别系统。

技术方案

一种融合音频分析和视频分析的欺诈识别方法，该方法包括：

接收当前用户的语音视频数据，从该语音视频数据中分离出音频及视频；

从音频中提取声音情绪特征，从视频中提取表情情绪特征；

将提取的声音情绪特征及表情情绪特征分别输入预先训练的语音识别模型和表情识别模型分别得到当前用户的情绪类别及评分分值；

根据预测精度对两个模型的识别结果分配权重，得到当前用户的欺诈识别结果。【依9月7日与毛工确认：每个模型识别出的是一种情绪及该情绪的评分，那么：1、加权融合得到的是一个什么结果？2.此处的结果应该是关于情绪的，还不能识别出是否有欺诈，那么如何识别是否存在欺诈？】

**从音频中提取声音情绪特征，包括以下步骤**：

第一特征提取步骤：从音频中提取大量的低阶音频特征，包括MFCC（Mel倒谱系数）语音特征、Pitch（音高）、Zero Crossing Rate（过零率）；

第二特征提取步骤：从上述音频低阶特征中提取动态回归系数，得到动态音频特征；

第三特征提取步骤：利用统计函数从所述低阶音频特征及动态音频特征中提取高阶音频特征；及

筛选步骤：利用特征筛选算法从所述高阶音频特征中筛选出最优特征子集。

**从视频中提取表情情绪特征，包括以下步骤**：

基础特征提取步骤：从视频中提取面部表情特征，包括面部动作（AU）、眼球朝向、头部朝向；

高阶特征构造步骤：统计各面部表情特征在视频中出现的次数及持续的时长，根据统计结果构造表情高阶特征；

筛选步骤：利用特征筛选算法对所述情高阶特征进行筛选出最优特征子集。

**所述语音识别模型及表情识别模型的训练，包括以下步骤**：

**视频切割步骤**：以情绪为单位对样本视频数据进行切割，得到一系列包含音频及视频的视频片段样本，给每个视频片段样本分配一个情绪标签；

**音频视频分离步骤**：从所述视频片段样本中分离出音频样本及视频样本；

**语音识别模型训练步骤**，包括：

第一特征提取步骤：从每个音频样本中提取大量的低阶音频特征，包括MFCC（Mel倒谱系数）语音特征、Pitch（音高）、Zero Crossing Rate（过零率）；【3、低阶音频特征，例如MFCC、Pitch、Zero Crossing Rate是什么格式？矩阵？向量？提供一个具体的示例】

第二特征提取步骤：从上述音频低阶特征中提取动态回归系数，得到动态音频特征； 【4、可否以具体的例子展示提取动态回归系数，从音频低阶特征得到的动态音频特征是什么格式？】

第三特征提取步骤：利用统计函数（如max-最大，min-最小, skewness-峰度， kurtosis-偏度）从所述低阶音频特征及动态音频特征中提取高阶音频特征；【5、结合3、4，以具体的例子展示从低阶音频特征及动态音频特征中提取的高阶音频特征】

筛选步骤：利用特征筛选算法从所述高阶音频特征中筛选出最优特征子集；

训练步骤：利用带情绪标注的各音频样本的最优特征子集对SVM进行训练得到语音识别模型，交叉验证确定语音识别模型的最优模型参数。

**表情识别模型训练步骤**，包括：

基础特征提取步骤：从每个视频片段中提取面部表情特征，包括面部动作（AU）、眼球朝向、头部朝向；

高阶特征构造步骤：统计各面部表情特征在每个视频片段中出现的次数及持续的时长，根据统计结果构造表情高阶特征；

筛选步骤：利用特征筛选算法对所述情高阶特征进行筛选出最优特征子集；

训练步骤：利用每个视频样本的最优特征子集对SVM进行训练得到表情识别模型，交叉验证确定表情识别模型的最优模型参数。