(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2020 年 6 月 25 日 (25.06.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/125397 A1

(51) 国际专利分类号:
G06K 9/00 (2006.01) *G06F 16/63* (2019.01)
G06K 9/62 (2006.01)

(21) 国际申请号: PCT/CN2019/122546

(22) 国际申请日: 2019 年 12 月 3 日 (03.12.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811546628.6 2018年12月18日 (18.12.2018) CN

(71) 申请人: 深圳壹账通智能科技有限公司 (ONE CONNECT SMART TECHNOLOGY CO.,LTD. (SHENZHEN)) [CN/CN]; 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518052 (CN)。

(72) 发明人: 江 波 (JIANG, Bo); 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518052 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区珠江东路 6 号 4501 房（部位：自编 01-03 和 08-12 单元）（仅限办公用途）, Guangdong 510623 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,

(54) **Title:** AUDIO DATA PUSHING METHOD AND DEVICE, COMPUTER APPARATUS, AND STORAGE MEDIUM

(54) **发明名称:** 音频数据推送方法、装置、计算机设备和存储介质

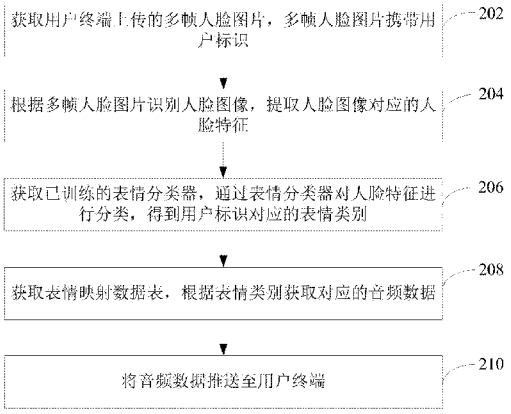

图 2

202 Obtain a plurality of face pictures uploaded by a user terminal, the plurality of face pictures carrying a user identifier
204 Identify a face image according to the plurality of face pictures, and extract a facial feature corresponding to the face image
206 Obtain a trained expression classifier, classify the facial feature by means of the expression classifier, and obtain an expression category corresponding to the user identifier
208 Obtain an expression mapping data table, and obtain corresponding audio data according to the expression category
210 Push the audio data to the user terminal

(57) **Abstract:** An audio data pushing method comprises: obtaining a plurality of face pictures uploaded by a user terminal, the plurality of face pictures carrying a user identifier; identifying a face image according to the plurality of face pictures, and extracting a facial feature corresponding to the face image; obtaining a trained expression classifier, classifying the facial feature by means of the expression classifier, and obtaining an expression category corresponding to the user identifier; obtaining an expression mapping data table, and obtaining corresponding audio data according to the expression category; and pushing the audio data to the user terminal.

(57) **摘要:** 一种音频数据推送方法, 包括: 获取用户终端上传的多帧人脸图片, 多帧人脸图片携带了用户标识; 根据多帧人脸图片识别人脸图像, 提取人脸图像对应的人脸特征; 获取已训练的表情分类器, 通过表情分类器对人脸特征进行分类, 得到用户标识对应的表情类别; 获取表情映射数据表, 根据表情类别获取对应的音频数据; 及将音频数据推送至用户终端.

WO 2020/125397 A1

[见续页]

JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

音频数据推送方法、装置、计算机设备和存储介质

相关申请的交叉引用：

本申请要求于 2018 年 12 月 18 日提交至中国专利局，申请号为 2018115466286 ，申请名称为“音频数据推送方法、装置、计算机设备和存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及一种音频数据推送方法、装置、计算机设备和存储介质。

背景技术

驾驶员在驾驶的过程中，行为受到一定限制。传统的车载播放器中的音乐更新较为繁琐，通常是随机为用户进行播放已下载的音乐或其他有声读物，当用户不喜欢当前的音乐时，需要手动更换当前的音乐，给用户带了不便。随着人工智能技术的迅速发展，出现了一些通过识别用户的语音指令执行相应的操作，但在驾驶过程中存在较大的噪音污染，语音指令识别的准确率较低，用户更换操作较为麻烦。

然而，发明人意识到，现有的一些音频数据自动推送的方式中，通常是根据排行或用户的历史记录相关的音频数据进行推送，而自动推送的音频播放数据不一定符合用户的喜好和情绪，导致了音乐或有声读物等音频播放数据的推送准确率较低。

发明内容

根据本申请公开的各种实施例，提供一种音频数据推送方法、装置、计算机设备和存储介质。

一种音频数据推送方法，所述方法包括：

获取用户终端上传的多帧人脸图片，所述多帧人脸图片携带了用户标识；

根据所述多帧人脸图片识别人脸图像，提取所述人脸图像对应的人脸特征；

获取已训练的表情分类器，通过所述表情分类器对所述人脸特征进行分类，得到所述用户标识对应的表情类别；

获取表情映射数据表，根据所述表情类别获取对应的音频数据；及

将所述音频数据推送至所述用户终端。

一种音频数据推送装置，所述装置包括：

数据获取模块，用于获取用户终端上传的多帧人脸图片，所述多帧人脸图片携带了用户标识；

人脸识别模块，用于根据所述多帧人脸图片识别人脸图像，提取所述人脸图像对应的人脸特征；

表情分类模块，用于获取已训练的表情分类器，通过所述表情分类器对所述人脸特征进行分类，得到所述用户标识对应的表情类别；

数据匹配模块，用于获取表情映射数据表，根据所述表情类别获取对应的音频数据；及

数据推送模块，用于将所述音频数据推送至所述用户终端。

一种计算机设备，包括存储器和一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述一个或多个处理器执行以下步骤：

获取用户终端上传的多帧人脸图片，所述多帧人脸图片携带了用户标识；

根据所述多帧人脸图片识别人脸图像，提取所述人脸图像对应的人脸特征；

获取已训练的表情分类器，通过所述表情分类器对所述人脸特征进行分类，得到所述用户标识对应的表情类别；

获取表情映射数据表，根据所述表情类别获取对应的音频数据；及

将所述音频数据推送至所述用户终端。

一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：

获取用户终端上传的多帧人脸图片，所述多帧人脸图片携带了用户标识；

根据所述多帧人脸图片识别人脸图像，提取所述人脸图像对应的人脸特征；

获取已训练的表情分类器，通过所述表情分类器对所述人脸特征进行分类，得到所述用户标识对应的表情类别；

获取表情映射数据表，根据所述表情类别获取对应的音频数据；及

将所述音频数据推送至所述用户终端。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其它的附图。

图 1 为根据一个或多个实施例中音频数据推送方法的应用场景图。

图 2 为根据一个或多个实施例中音频数据推送方法的流程示意图。

图 3 为根据一个或多个实施例中训练表情分类器步骤的流程示意图。

图 4 为根据一个或多个实施例中通过表情分类器对人脸特征分类步骤的流程示意图。

图 5 为根据一个或多个实施例中音频数据推送装置的框图；

图 6 为根据一个或多个实施例中计算机设备的框图。

具体实施方式

为了使本申请的技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

本申请提供的音频数据推送方法，可以应用于如图 1 所示的应用环境中。用户终端 102 通过网络与服务器 104 通过网络进行通信。用户终端 102 可以是安装在车辆中具有摄像功能的车载系统终端，也可以是与车辆分离的具有摄像功能的智能手机、平板电脑和便携式设备等，服务器 104 可以用独立的服务器或者是多个服务器组成的服务器集群来实现。服务器 104 获取用户终端 102 上传的多帧人脸图片，多帧人脸图片携带了用户标识。服务器 104 进而根据多帧人脸图片识别人脸图像，并提取人脸图像对应的人脸特征。服务器 104 则进一步获取预设的表情分类器，通过表情分类器对人脸特征进行分类，得到用户标识对应的表情类别。通过利用基于神经网络模型的表情分类器对用户的多帧人脸特征进行识别并分类，由此能够准确有效地识别出该用户当前的表情类别。服务器 104 识别出该用户当前的人脸特征对应的表情类别后，则获取预设的表情映射数据表，根据表情类别获取对应的音频数据，并将音频数据推送至用户终端 102。通过对用户在驾驶过程中的人脸表情特征进行识别，并通过基于神经网络的表情分类器对用户的表情进行识别分类，能够准确有效地识别出用户当前的表情情绪，进而能够有效地根据用户的表情情绪推送相对应的音频数据，由此能够有效地提高音频数据的推送准确率。

在其中一个实施例中，如图 2 所示，提供了一种音频数据推送方法，以该方法应用于图 1 中的服务器为例进行说明，包括以下步骤：

步骤 202，获取用户终端上传的多帧人脸图片，多帧人脸图片携带用户标识。

用户终端可以是安装在车辆中具有摄像功能的车载系统终端，也可以是与车辆分离的具有摄像功能的智能手机、平板电脑和便携式设备等。用户在驾驶的过程中，用户开启对应的用户终端后，可以向用户终端触发播放推荐指令。用户终端响应播放推荐指令后，通过摄像装置按照预设频率捕捉用户的人脸图片，并将捕捉的多帧人脸图片按照时间序列上传至服务器。其中，车载终端捕捉的多帧人脸图片可以是连续的多帧静态图片，也可以是动态视频。

进一步地，用户开启用户终端中相应的应用程序后，用户终端还可以根据预设的频率通过摄像装置自动拍摄用户的人脸图片或人脸视频，并将拍摄的人脸图片或人脸视频按照时间序列上传至服务器。

服务器进而获取用户终端上传的通过摄像装置拍摄的多帧人脸图片，多帧人脸图片携带了用户标识。

步骤 204，根据多帧人脸图片识别人脸图像，提取人脸图像对应的人脸特征。

服务器获取用户终端通过摄像装置拍摄的多帧人脸图片后，则对多帧人脸图片进行人脸识别，识别出多帧人脸图片中的人脸图像，并对识别出的多帧人脸图像进行特征提取，提取出每帧人脸图像对应的人脸特征。多帧人脸图片还包括对应的图像序列。具体地，服务器对多帧人脸图片进行关键点定位，根据预设的人脸识别算法识别定位后的多帧人脸图片中的多帧人脸图像，并对多帧人脸图像进行特征提取，得到多帧人脸图像对应的人脸特征，由此能够有效地识别并提取出人脸图像对应的人脸特征。

步骤 206，获取已训练的表情分类器，通过表情分类器对人脸特征进行分类，得到用户标识对应的表情类别。

服务器提取每帧人脸图像对应的人脸特征后，获取预设的表情分类器。表情分类器可以是预先训练得到的基于神经网络模型的分类器。具体地，服务器将提取的每帧人脸图像对应的人脸特征输入至表情分类器中，通过表情分类器中的卷积神经网络识别每帧人脸特征，并根据图像序列和多帧人脸特征计算对应的动态人脸特征，进而计算动态人脸特征属于每个表情类别的概率值。服务器则获取概率值最高的表情类别，由此得到对用户标识对应的表情类别。

步骤 208，获取预设的表情映射数据表，根据表情类别获取对应的音频数据。

步骤 210，将音频数据推送至用户终端。

服务器通过表情分类器对多帧人脸特征进行分类，得到对应的表情类别后，进一步获取预设的表情映射数据表。其中，表情映射数据表中包括表情类别和对应的心情标签，以及心

情标签对应的音频数据标识。具体地，服务器可以预先获取多个视频数据，每个视频数据可以包括对应的视频数据标识。服务器进一步获取预设的视频数据分类模型，通过视频数据分类模型对多个视频数据进行分类，并对视频数据添加对应的心情标签。心情标签可以对应一个或多个表情类别。服务器则根据心情标签获取对应的表情类别，并根据心情标签和表情类别以及对应的音频数据标识建立表情映射数据表。

服务器则根据表情映射数据表匹配与表情类别对应的心情标签，并根据心情标签获取对应的音频数据。其中，音频数据可以包括各种音乐、广播音频以及有声读物等数据。

例如，心情标签可以对应音乐播放列表，服务器识别出该用户当前的表情类别后，则将表情类别对应的播放列表中的音频数据推送给用户终端，使得用户终端对推送的音频数据进行播放。通过对用户在驾驶过程中的人脸表情特征进行识别，并通过基于神经网络的表情分类器对用户的表情进行识别分类，能够准确有效地识别出用户当前的表情情绪，进而能够有效地根据用户的表情情绪推送相对应的音频数据，由此能够有效地提高音频数据的推送准确率。

上述音频数据推送方法中，服务器获取用户终端上传的多帧人脸图片，多帧人脸图片携带了用户标识。服务器进而根据多帧人脸图片识别人脸图像，并提取人脸图像对应的人脸特征。服务器则进一步获取预设的表情分类器，通过表情分类器对人脸特征进行分类，得到用户标识对应的表情类别。通过利用基于神经网络模型的表情分类器对用户的多帧人脸特征进行识别并分类，由此能够准确有效地识别出该用户当前的表情类别。服务器识别出该用户当前的人脸特征对应的表情类别后，则获取预设的表情映射数据表，根据表情类别获取对应的音频数据，并将音频数据推送至用户终端。通过对用户在驾驶过程中的人脸表情特征进行识别，并通过基于神经网络的表情分类器对用户的表情进行识别分类，能够准确有效地识别出用户当前的表情情绪，进而能够有效地根据用户的表情情绪推送相对应的音频数据，由此能够有效地提高音频数据的推送准确率。

在其中一个实施例中，根据多帧人脸图片识别人脸图像，提取人脸图像对应的人脸特征的步骤，包括：对多帧人脸图片进行关键点定位，得到定位后的多帧人脸图片；根据预设人脸识别算法识别定位后的多帧人脸图片中的多帧人脸图像，得到多帧人脸图像对应的人脸特征。

用户终端可以根据预设的频率通过摄像装置捕捉用户的多帧人脸图片，并将捕捉的多帧人脸图片上传至服务器，多帧人脸图片携带了用户标识。服务器获取用户终端上传的多帧人脸图片后，进而根据多帧人脸图片识别人脸图像，并提取人脸图像对应的人脸特征。

具体地，服务器根据预设算法检测多帧人脸图片中的关键点，并对多帧人脸图片进行关键点定位。服务器进而根据预设的人脸识别算法识别关键点定位后的多帧人脸图片中的多帧人脸图像，服务器还可以对人脸图片进行分割，提取出人脸图像部分，进一步对提取出的人脸图像进行灰度化处理和归一化处理。服务器则对多帧人脸图像进行特征提取，得到多帧人脸图像对应的人脸特征，由此能够有效地识别并提取出人脸图像对应的人脸特征。

服务器还可以通过识别每帧人脸图片中的人脸特征是否一致，来判断用户的的人脸表情开始帧和结束帧以及持续的时间，通过人脸表情开始帧和结束帧以及持续的时间提取出用户的当前的多帧人脸特征。由此能够准确有效地提取出用户当前表情对应的多帧人脸特征。

服务器识别并提取出该用户的多帧人脸特征后，进一步获取预设的表情分类器，通过表情分类器对人脸特征进行分类，得到用户标识对应的表情类别。通过利用基于神经网络模型的表情分类器对用户的多帧人脸特征进行识别并分类，由此能够准确有效地识别出该用户当前的表情类别。服务器识别出该用户当前的人脸特征对应的表情类别后，则获取预设的表情映射数据表，根据表情类别获取对应的音频数据，并将音频数据推送至用户终端。通过对用户在驾驶过程中的人脸表情特征进行识别，能够准确有效地提取出用户当前表情对应的多帧人脸特征。并通过基于神经网络的表情分类器对用户的表情进行识别分类，能够准确有效地识别出用户当前的表情情绪，进而能够有效地根据用户的表情情绪推送相对应的音频数据，由此能够有效地提高音频数据的推送准确率。

在其中一个实施例中，如图 3 所示，在获取预设的表情分类器之前，该方法还包括训练表情分类器的步骤，该步骤具体包括以下内容：

步骤 302，从预设数据库中获取多个表情数据。

步骤 304，利用获取的多个表情数据生成训练集和验证集；训练集中包括已标注的表情数据，验证集中包括未标注的表情数据。

步骤 306，利用训练集中已标注的表情数据通过预设算法进行训练得到初步的表情分类器。

步骤 308，将验证集中未标注的表情数据输入至初步的表情分类器中进行验证训练。

步骤 310，直到达到预设概率值的验证集数据的数量达到预设比值时，则停止训练，得到训练完成的表情分类器。

服务器在获取预设的表情分类器之前，还需要利用大量的表情数据训练得到表情分类器。具体地，服务器可以从本地或第三方数据库中获取大量的表情数据，表情数据可以包括表情图片和表情视频以及动态表情图像等。服务器进而将获取的大量的表情数据生成训练集

和验证集，其中，训练集中的表情数据可以是通过人工标注后的表情数据，验证集中的表情数据可以是未进行标注的表情数据。

服务器则将训练集中的表情数据进行特征提取，通过预设的算法进行训练，训练得到初步的表情分类器。例如，服务器可以通过 CNN（Convolutional Neural Network，卷积神经网络）对表情数据中的人脸图像进行卷积操作，得到人脸图像对应的人脸特征。并将表情数据中多帧人脸图片对应的多帧人脸特征输入至 BLSTM（Bidirectional Long Short-term Memory，双向长短期记忆神经网络），通过预设函数计算出表情数据对应的动态表情特征。服务器进而根据预设算法计算出每个动态表情特征对应每个标签类别的概率值，从而对表情分类器进行训练，由此能够有效地得到初步的表情分类器。

服务器进一步将验证集中的数据输入至初始表情分类器中进行持续训练，得到每个表情数据对应每个类别的概率值，获取达到预设的概率阈值的表情数据，当达到预设的概率阈值的表情数据的数量达到预设比值时，则停止训练，则得到训练完成的表情分类器。通过利用大量的表情数据对表情分类器进行训练和验证，从而可以有效地训练出分类准确率较高的表情分类器。

在其中一个实施例中，如图 4 所示，多帧人脸图片包括对应的图像序列，通过表情分类器对人脸特征进行分类，得到用户标识对应的表情类别的步骤，具体包括一下内容：

步骤 402，将多帧人脸图片对应的多帧人脸特征输入至表情分类器，通过表情分类器中的卷积神经网络识别每帧人脸特征向量。

步骤 404，根据图像序列和多帧人脸特征向量计算对应的动态表情特征。

步骤 406，计算动态表情特征属于每个表情类别的概率值。

步骤 408，获取概率值最高的表情类别，得到用户标识对应的表情类别。

用户终端可以根据预设的频率通过摄像装置捕捉用户的多帧人脸图片，并将捕捉的多帧人脸图片上传至服务器，多帧人脸图片携带了用户标识。服务器获取用户终端上传的多帧人脸图片后，进而根据多帧人脸图片识别人脸图像，并提取人脸图像对应的人脸特征。

服务器提取每帧人脸图像对应的人脸特征后，获取预设的表情分类器。其中，表情分类器可以是预先训练得到的基于神经网络模型的分类器。具体地，服务器将提取的每帧人脸图像对应的人脸特征输入至表情分类器中，通过表情分类器中的卷积神经网络识别每帧人脸特征向量，并根据图像序列和多帧人脸特征向量计算对应的动态人脸特征，进而计算动态人脸特征属于每个表情类别的概率值。服务器则获取概率值最高的表情类别，由此得到对用户标识对应的表情类别。

例如，服务器可以通过 CNN（Convolutional Neural Network，卷积神经网络）对多帧人脸图像对应的多帧人脸特征进行卷积操作，识别得到人脸图像对应的人脸特征向量。并将多帧人脸特征向量输入至 BLSTM（Bidirectional Long Short-term Memory，双向长短期记忆神经网络），通过预设函数计算出多帧人脸特征向量对应的动态表情特征。服务器进而根据预设算法计算出每个动态表情特征对应每个标签类别的概率值，服务器则获取概率值最高的表情类别。

服务器识别出该用户当前的人脸特征对应的表情类别后，则获取预设的表情映射数据表，根据表情类别获取对应的音频数据，并将音频数据推送至用户终端。通过基于神经网络的表情分类器对用户的表情进行识别分类，能够准确有效地识别出用户当前的表情情绪，进而能够有效地根据用户的表情情绪推送相对应的音频数据，由此能够有效地提高音频数据的推送准确率。

在其中一个实施例中，获取表情映射数据表之前，还包括：获取多个视频数据；获取已训练的视频数据分类模型，通过视频数据分类模型对多个视频数据进行分类，根据分类结果对多个视频数据添加对应的心情标签；根据心情标签获取对应的表情类别，心情标签对应一个或多个表情类别；根据心情标签和表情类别以及对应的音频数据建立表情映射数据表。

服务器在获取预设的表情数据映射表之前，还可以预先建立表情数据映射表。具体地，服务器可以预先获取多个视频数据，每个视频数据可以包括对应的视频数据标识。服务器进一步获取预设的视频数据分类模型，通过视频数据分类模型对多个视频数据进行分类，根据分类结果对多个视频数据添加对应的心情标签。心情标签可以对应一个或多个表情类别。服务器则根据心情标签获取对应的表情类别，并根据心情标签和表情类别以及对应的音频数据标识建立表情映射数据表，由此能够有效地建立用户表情与音频数据之间的关联关系。

服务器建立表情映射数据表后，服务器获取用户终端上传的多帧人脸图片，进而根据多帧人脸图片识别人脸图像，并提取人脸图像对应的人脸特征。服务器获取预设的表情分类器，通过表情分类器对多帧人脸特征进行分类，得到对应的表情类别后，进一步获取预设的表情映射数据表。其中，表情映射数据表中包括表情类别和对应的心情标签，以及心情标签对应的音频数据标识。

服务器则根据表情映射数据表匹配与表情类别对应的心情标签，并根据心情标签获取对应的音频数据。音频数据可以包括各种音乐、广播音频以及有声读物等数据。服务器则将获取的音频数据推送至用户终端。

例如，心情标签可以对应音乐播放列表，服务器识别出该用户当前的表情类别后，则将

表情类别对应的播放列表中的音频数据推送给用户终端，使得用户终端对推送的音频数据进行播放。通过对用户在驾驶过程中的人脸表情特征进行识别，并通过基于神经网络的表情分类器对用户的表情进行识别分类，能够准确有效地识别出用户当前的表情情绪，进而能够有效地根据用户的表情情绪推送相对应的音频数据，由此能够有效地提高音频数据的推送准确率。

在其中一个实施例中，该方法还包括：根据预设的频率获取用户标识对应的历史记录数据；获取预设的分析模型，通过分析模型对历史数据进行分析，得到分析结果；根据表情类别标签和分析结果匹配对应的音频数据；获取匹配度最高的音频数据，将音频数据推送至用户终端。

服务器还可以根据预设的频率获取用户的历史记录数据，例如用户的听歌记录和点播记录等。服务器通过对用户的历史记录数据进行大数据分析，具体地，服务器可以获取预设的分析模型，分析模型可以是基于神经网络的模型，也可以是基于决策树的模型。服务器将用户的历史记录数据输入至分析模型中，通过分析模型对历史数据进行分析，并得到对应的得到分析结果。

用户终端可以根据预设的频率通过摄像装置捕捉用户的多帧人脸图片，并将捕捉的多帧人脸图片上传至服务器，多帧人脸图片携带了用户标识。服务器获取用户终端上传的多帧人脸图片后，进而根据多帧人脸图片识别人脸图像，并提取人脸图像对应的人脸特征。

服务器识别并提取出该用户的多帧人脸特征后，进一步获取预设的表情分类器，通过表情分类器对人脸特征进行分类，得到用户标识对应的表情类别。通过利用基于神经网络模型的表情分类器对用户的多帧人脸特征进行识别并分类，由此能够准确有效地识别出该用户当前的表情类别。

服务器识别出该用户当前的人脸特征对应的表情类别后，则获取该用户标识对应的历史数据的分析结果。服务器则根据表情类别标签和分析结果匹配对应的音频数据。具体地，服务器获取预设的表情映射数据表，根据表情类别匹配对应的音频数据，并根据分析结果获取与分析结果相匹配的音频数据。服务器进而将音频数据推送至用户终端。例如，服务器可以根据用户的常听记录和收藏记录等数据分析出用户对音频数据的偏好度，并根据用户当前的表情像用户终端推送该标签类别对应的用户偏好度较高的音频数据。通过对用户在驾驶过程中的人脸表情特征进行识别，能够准确有效地提取出用户当前表情对应的多帧人脸特征。通过基于神经网络的表情分类器对用户的表情进行识别分类，能够准确有效地识别出用户当前的表情类别。并通过对用户的历史数据进行分析，根据用户的表情和偏好对用户进行推送音

频数据，进而能够有效地根据用户的表情推送相用户喜好的音频数据，由此能够有效地提高音频数据的推送准确率。

在其中一个实施例中，该方法还包括：当表情类别标签为疲劳时，获取对应的提示信息和音频数据；将提示信息和音频数据发送至用户终端，以使用户终端按照预设的方式进行提示和播放音频数据。

用户终端可以根据预设的频率通过摄像装置捕捉用户的多帧人脸图片，并将捕捉的多帧人脸图片上传至服务器，多帧人脸图片携带了用户标识。服务器获取用户终端上传的多帧人脸图片后，进而根据多帧人脸图片识别人脸图像，并提取人脸图像对应的人脸特征。

服务器识别并提取出该用户的多帧人脸特征后，进一步获取预设的表情分类器，通过表情分类器对人脸特征进行分类，得到用户标识对应的表情类别。通过利用基于神经网络模型的表情分类器对用户的多帧人脸特征进行识别并分类，由此能够准确有效地识别出该用户当前的表情类别。

服务器识别出该用户当前的人脸特征对应的表情类别后，当服务器识别出用户当前的表情类别为疲劳时，获取相对应的提示信息和音频数据，并将提示信息和音频数据发送至用户终端，以使用户终端按照预设的方式进行提示和播放音频数据。用户在驾驶过程中，若用户出现疲劳驾驶时，此时存在较高的风险隐患，服务器通过表情识别出用户的当前存在疲劳现象时，按照预设方式进行加强提醒，如高分贝音量的音频和提示语音进行提醒，由此能够有效地对用户进行提醒，以提示用户安全驾驶。

应该理解的是，虽然图 2-4 的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，图 2-4 中的至少一部分步骤可以包括多个子步骤或者多个阶段，这些子步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些子步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

在其中一个实施例中，如图 5 所示，提供了一种音频数据推送装置，包括：数据获取模块 502、人脸识别模块 504、表情分类模块 506、数据匹配模块 508 和数据推送模块 510，其中：

数据获取模块 502，用于获取用户终端上传的多帧人脸图片，多帧人脸图片携带了用户标识；

人脸识别模块 504，用于根据多帧人脸图片识别人脸图像，提取人脸图像对应的人脸特

征；

表情分类模块 506，用于获取已训练的表情分类器，通过表情分类器对人脸特征进行分类，得到用户标识对应的表情类别；

数据匹配模块 508，用于获取表情映射数据表，根据表情类别获取对应的音频数据；

数据推送模块 510，用于将音频数据推送至用户终端。

在其中一个实施例中，人脸识别模块 504 还用于对多帧人脸图片进行关键点定位，得到定位后的多帧人脸图片；根据预设的人脸识别算法识别定位后的多帧人脸图片中的多帧人脸图像，得到多帧人脸图像对应的人脸特征。

在其中一个实施例中，该装置还包括表情分类器训练模块，用于从预设数据库中获取多个表情数据；利用获取的多个表情数据生成训练集和验证集；训练集中包括已标注的表情数据，验证集中包括未标注的表情数据；利用训练集中已标注的表情通过预设算法进行训练得到初步的表情分类器；将验证集中未标注的表情数据输入至初步的表情分类器中进行验证训练；直到达到预设概率值的验证集数据的数量达到预设比值时，则停止训练，得到训练完成的表情分类器。

在其中一个实施例中，多帧人脸图片包括对应的图像序列，表情分类模块 506 还用于将多帧人脸图片对应的多帧人脸特征输入至表情分类器，通过表情分类器中的卷积神经网络识别每帧人脸特征向量；根据图像序列和多帧人脸特征向量计算对应的动态表情特征；计算动态表情特征属于每个表情类别的概率值；获取概率值最高的表情类别，得到用户标识对应的表情类别。

在其中一个实施例中，该装置还包括表情映射数据表建立模块，用于获取多个视频数据；获取已训练的视频数据分类模型，通过视频数据分类模型对多个视频数据进行分类，根据分类结果对多个视频数据添加对应的心情标签；根据心情标签获取对应的表情类别，心情标签对应一个或多个表情类别；根据心情标签和表情类别以及对应的音频数据建立表情映射数据表。

在其中一个实施例中，该装置还包括数据分析模块，用于根据预设频率获取用户标识对应的历史记录数据；获取预设的分析模型，通过分析模型对历史数据进行分析，得到分析结果；数据推送模块 510 还用于根据表情类别标签和分析结果匹配对应的音频数据；获取相匹配的音频数据，将音频数据推送至用户终端。

在其中一个实施例中，该装置还包括提示模块，用于当表情类别标签为疲劳时，获取对应的提示信息和音频数据；将提示信息和音频数据发送至用户终端，以使用户终端按照预设的方

式进行提示和播放音频数据。

关于音频数据推送装置的具体限定可以参见上文中对于音频数据推送方法的限定，在此不再赘述。上述音频数据推送装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 6 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储人脸图片、表情类别和音频数据等数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种音频数据推送方法。

本领域技术人员可以理解，图 6 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

一种计算机设备，包括存储器和一个或多个处理器，存储器中储存有计算机可读指令，计算机可读指令被处理器执行时，使得一个或多个处理器执行以下步骤：

获取用户终端上传的多帧人脸图片，多帧人脸图片携带了用户标识；

根据多帧人脸图片识别人脸图像，提取人脸图像对应的人脸特征；

获取已训练的表情分类器，通过表情分类器对人脸特征进行分类，得到用户标识对应的表情类别；

获取表情映射数据表，根据表情类别获取对应的音频数据；及

将音频数据推送至用户终端。

在一个实施例中，处理器执行计算机可读指令时还实现以下步骤：对多帧人脸图片进行关键点定位，得到定位后的多帧人脸图片；及根据预设人脸识别算法识别定位后的多帧人脸图片中的多帧人脸图像，得到多帧人脸图像对应的人脸特征。

在一个实施例中，处理器执行计算机可读指令时还实现以下步骤：从预设数据库中获取多个表情数据；利用获取的多个表情数据生成训练集和验证集；训练集中包括已标注的表情数据，验证集中包括未标注的表情数据；利用训练集中已标注的表情数据通过预设算法进行

训练得到初步的表情分类器；将验证集中未标注的表情数据输入至初步的表情分类器中进行验证训练；及直到达到预设概率值的验证集数据的数量达到预设比值时，则停止训练，得到训练完成的表情分类器。

在一个实施例中，多帧人脸图片包括对应的图像序列，处理器执行计算机可读指令时还实现以下步骤：将多帧人脸图片对应的多帧人脸特征输入至表情分类器，通过表情分类器中的卷积神经网络识别每帧人脸特征向量；根据图像序列和多帧人脸特征向量计算对应的动态表情特征；计算动态表情特征属于每个表情类别的概率值；及获取概率值最高的表情类别，得到用户标识对应的表情类别。

在一个实施例中，处理器执行计算机可读指令时还实现以下步骤：获取多个视频数据；获取已训练的视频数据分类模型，通过视频数据分类模型对多个视频数据进行分类，根据分类结果对多个视频数据添加对应的心情标签；根据心情标签获取对应的表情类别，心情标签对应一个或多个表情类别；及根据心情标签和表情类别以及对应的音频数据建立表情映射数据表。

在一个实施例中，处理器执行计算机可读指令时还实现以下步骤：根据预设频率获取用户标识对应的历史记录数据；获取预设的分析模型，通过分析模型对历史数据进行分析，得到分析结果；根据表情类别标签和分析结果匹配对应的音频数据；及获取相匹配的音频数据，将音频数据推送至用户终端。

在一个实施例中，处理器执行计算机可读指令时还实现以下步骤：当表情类别标签为疲劳时，获取对应的提示信息和音频数据；及将提示信息和音频数据发送至用户终端，以使用户终端按照预设的方式进行提示和播放音频数据。

一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：获取用户终端上传的多帧人脸图片，多帧人脸图片携带了用户标识；

根据多帧人脸图片识别人脸图像，提取人脸图像对应的人脸特征；

获取已训练的表情分类器，通过表情分类器对人脸特征进行分类，得到用户标识对应的表情类别；

获取表情映射数据表，根据表情类别获取对应的音频数据；

将音频数据推送至用户终端。

在一个实施例中，计算机可读指令被处理器执行时还实现以下步骤：对多帧人脸图片进行关键点定位，得到定位后的多帧人脸图片；及根据预设人脸识别算法识别定位后的多帧人

脸图片中的多帧人脸图像，得到多帧人脸图像对应的人脸特征。

在一个实施例中，计算机可读指令被处理器执行时还实现以下步骤：从预设数据库中获取多个表情数据；利用获取的多个表情数据生成训练集和验证集；训练集中包括已标注的表情数据，验证集中包括未标注的表情数据；利用训练集中已标注的表情数据通过预设算法进行训练得到初步的表情分类器；将验证集中未标注的表情数据输入至初步的表情分类器中进行验证训练；及直到达到预设概率值的验证集数据的数量达到预设比值时，则停止训练，得到训练完成的表情分类器。

在一个实施例中，多帧人脸图片包括对应的图像序列，计算机可读指令被处理器执行时还实现以下步骤：将多帧人脸图片对应的多帧人脸特征输入至表情分类器，通过表情分类器中的卷积神经网络识别每帧人脸特征向量；根据图像序列和多帧人脸特征向量计算对应的动态表情特征；计算动态表情特征属于每个表情类别的概率值；及获取概率值最高的表情类别，得到用户标识对应的表情类别。

在一个实施例中，计算机可读指令被处理器执行时还实现以下步骤：获取多个视频数据；获取已训练的视频数据分类模型，通过视频数据分类模型对多个视频数据进行分类，根据分类结果对多个视频数据添加对应的心情标签；根据心情标签获取对应的表情类别，心情标签对应一个或多个表情类别；及根据心情标签和表情类别以及对应的音频数据建立表情映射数据表。

在一个实施例中，计算机可读指令被处理器执行时还实现以下步骤：根据预设频率获取用户标识对应的历史记录数据；获取预设的分析模型，通过分析模型对历史数据进行分析，得到分析结果；根据表情类别标签和分析结果匹配对应的音频数据；及获取相匹配的音频数据，将音频数据推送至用户终端。

在一个实施例中，计算机可读指令被处理器执行时还实现以下步骤：当表情类别标签为疲劳时，获取对应的提示信息和音频数据；及将提示信息和音频数据发送至用户终端，以使用户终端按照预设的方式进行提示和播放音频数据。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读取存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程ROM（PROM）、电可编程ROM（EPROM）、电可擦除可编程ROM（EEPROM）或闪存。

易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM 以多种形式可得，诸如静态 RAM（SRAM）、动态 RAM（DRAM）、同步 DRAM（SDRAM）、双数据率 SDRAM（DDRSDRAM）、增强型 SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接 RAM（RDRAM）、直接存储器总线动态 RAM（DRDRAM）、以及存储器总线动态 RAM（RDRAM）等。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种音频数据推送方法，所述方法包括：

获取用户终端上传的多帧人脸图片，所述多帧人脸图片携带了用户标识；

根据所述多帧人脸图片识别人脸图像，提取所述人脸图像对应的人脸特征；

获取已训练的表情分类器，通过所述表情分类器对所述人脸特征进行分类，得到所述用户标识对应的表情类别；

获取表情映射数据表，根据所述表情类别获取对应的音频数据；及

将所述音频数据推送至所述用户终端。

2、根据权利要求 1 所述的方法，其特征在于，所述根据所述多帧人脸图片识别人脸图像，提取所述人脸图像对应的人脸特征，包括：

对所述多帧人脸图片进行关键点定位，得到定位后的多帧人脸图片；及

根据预设人脸识别算法识别定位后的多帧人脸图片中的多帧人脸图像，得到多帧人脸图像对应的人脸特征。

3、根据权利要求 1 所述的方法，其特征在于，所述获取预设的表情分类器之前，所述方法还包括：

从预设数据库中获取多个表情数据；

利用获取的多个表情数据生成训练集和验证集；所述训练集中包括已标注的表情数据，所述验证集中包括未标注的表情数据；

利用所述训练集中已标注的表情数据通过预设算法进行训练得到初步的表情分类器；

将所述验证集中未标注的表情数据输入至所述初步的表情分类器中进行验证训练；及

直到达到预设概率值的验证集数据的数量达到预设比值时，则停止训练，得到训练完成的表情分类器。

4、根据权利要求 1 至 3 任意一项所述的方法，其特征在于，所述多帧人脸图片包括对应的图像序列，所述通过所述表情分类器对所述人脸特征进行分类，得到所述用户标识对应的表情类别，包括：

将多帧人脸图片对应的多帧人脸特征输入至所述表情分类器，通过所述表情分类器中的卷积神经网络识别每帧人脸特征向量；

根据所述图像序列和多帧人脸特征向量计算对应的动态表情特征；

计算所述动态表情特征属于每个表情类别的概率值；及

获取所述概率值最高的表情类别，得到所述用户标识对应的表情类别。

5、根据权利要求 1 所述的方法，其特征在于，所述获取表情映射数据表之前，还包括：

获取多个视频数据；

获取已训练的视频数据分类模型，通过所述视频数据分类模型对多个视频数据进行分类，根据分类结果对多个视频数据添加对应的心情标签；

根据所述心情标签获取对应的表情类别，所述心情标签对应一个或多个表情类别；及

根据所述心情标签和所述表情类别以及对应的音频数据建立表情映射数据表。

6、根据权利要求 1 所述的方法，其特征在于，所述方法还包括：

根据预设频率获取所述用户标识对应的历史记录数据；

获取预设的分析模型，通过所述分析模型对所述历史数据进行分析，得到分析结果；

根据所述表情类别标签和所述分析结果匹配对应的音频数据；及

获取相匹配的音频数据，将所述音频数据推送至所述用户终端。

7、根据权利要求 1 所述的方法，其特征在于，所述方法还包括：

当所述表情类别标签为疲劳时，获取对应的提示信息和音频数据；及

将所述提示信息和音频数据发送至所述用户终端，以使所述用户终端按照预设的方式进行提示和播放所述音频数据。

8、一种音频数据推送装置，所述装置包括：

数据获取模块，用于获取用户终端上传的多帧人脸图片，所述多帧人脸图片携带了用户标识；

人脸识别模块，用于根据所述多帧人脸图片识别人脸图像，提取所述人脸图像对应的人脸特征；

表情分类模块，用于获取已训练的表情分类器，通过所述表情分类器对所述人脸特征进行分类，得到所述用户标识对应的表情类别；

数据匹配模块，用于获取预设的表情映射数据表，根据所述表情类别获取对应的音频数据；及

数据推送模块，用于将所述音频数据推送至所述用户终端。

9、根据权利要求 8 所述的装置，其特征在于，所述表情分类模块还用于将多帧人脸图片对应的多帧人脸特征输入至所述表情分类器，通过所述表情分类器中的卷积神经网络识别每帧人脸特征向量；根据所述图像序列和多帧人脸特征向量计算对应的动态表情特征；计算所述动态表情特征属于每个表情类别的概率值；及获取所述概率值最高的表情类别，得到所述用户标识对应的表情类别。

10、根据权利要求 8 所述的装置，其特征在于，所述数据匹配模块还用于根据预设频率获取所述用户标识对应的历史记录数据；获取预设的分析模型，通过所述分析模型对所述历史数据进行分析，得到分析结果；根据所述表情类别标签和所述分析结果匹配对应的音频数据；及获取相匹配的音频数据，将所述音频数据推送至所述用户终端。

11、一种计算机设备，包括存储器及一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

获取用户终端上传的多帧人脸图片，所述多帧人脸图片携带了用户标识；

根据所述多帧人脸图片识别人脸图像，提取所述人脸图像对应的人脸特征；

获取已训练的表情分类器，通过所述表情分类器对所述人脸特征进行分类，得到所述用户标识对应的表情类别；

获取表情映射数据表，根据所述表情类别获取对应的音频数据；及

将所述音频数据推送至所述用户终端。

12、根据权利要求 11 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时还执行以下步骤：从预设数据库中获取多个表情数据；利用获取的多个表情数据生成训练集和验证集；所述训练集中包括已标注的表情数据，所述验证集中包括未标注的表情数据；利用所述训练集中已标注的表情数据通过预设算法进行训练得到初步的表情分类器；将所述验证集中未标注的表情数据输入至所述初步的表情分类器中进行验证训练；及直到达到预设概率值的验证集数据的数量达到预设比值时，则停止训练，得到训练完成的表情分类器。

13、根据权利要求 11 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时还执行以下步骤：将多帧人脸图片对应的多帧人脸特征输入至所述表情分类器，通过所述表情分类器中的卷积神经网络识别每帧人脸特征向量；根据所述图像序列和多帧人脸特征向量计算对应的动态表情特征；计算所述动态表情特征属于每个表情类别的概率值；及获取所述概率值最高的表情类别，得到所述用户标识对应的表情类别。

14、根据权利要求 11 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时还执行以下步骤：获取多个视频数据；获取已训练的视频数据分类模型，通过所述视频数据分类模型对多个视频数据进行分类，根据分类结果对多个视频数据添加对应的心情标签；根据所述心情标签获取对应的表情类别，所述心情标签对应一个或多个表情类别；及根据所述心情标签和所述表情类别以及对应的音频数据建立表情映射数据表。

15、根据权利要求 11 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读

指令时还执行以下步骤：根据预设频率获取所述用户标识对应的历史记录数据；获取预设的分析模型，通过所述分析模型对所述历史数据进行分析，得到分析结果；根据所述表情类别标签和所述分析结果匹配对应的音频数据；及获取相匹配的音频数据，将所述音频数据推送至所述用户终端。

16、一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：

获取用户终端上传的多帧人脸图片，所述多帧人脸图片携带了用户标识；

根据所述多帧人脸图片识别人脸图像，提取所述人脸图像对应的人脸特征；

获取已训练的表情分类器，通过所述表情分类器对所述人脸特征进行分类，得到所述用户标识对应的表情类别；

获取表情映射数据表，根据所述表情类别获取对应的音频数据；及

将所述音频数据推送至所述用户终端。

17、根据权利要求 16 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：从预设数据库中获取多个表情数据；利用获取的多个表情数据生成训练集和验证集；所述训练集中包括已标注的表情数据，所述验证集中包括未标注的表情数据；利用所述训练集中已标注的表情数据通过预设算法进行训练得到初步的表情分类器；将所述验证集中未标注的表情数据输入至所述初步的表情分类器中进行验证训练；及直到达到预设概率值的验证集数据的数量达到预设比值时，则停止训练，得到训练完成的表情分类器。

18、根据权利要求 16 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：将多帧人脸图片对应的多帧人脸特征输入至所述表情分类器，通过所述表情分类器中的卷积神经网络识别每帧人脸特征向量；根据所述图像序列和多帧人脸特征向量计算对应的动态表情特征；计算所述动态表情特征属于每个表情类别的概率值；及获取所述概率值最高的表情类别，得到所述用户标识对应的表情类别。

19、根据权利要求 16 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：获取多个视频数据；获取已训练的视频数据分类模型，通过所述视频数据分类模型对多个视频数据进行分类，根据分类结果对多个视频数据添加对应的心情标签；根据所述心情标签获取对应的表情类别，所述心情标签对应一个或多个表情类别；及根据所述心情标签和所述表情类别以及对应的音频数据建立表情映射数据表。

20、根据权利要求 16 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：根据预设频率获取所述用户标识对应的历史记录数据；获取预设的

分析模型，通过所述分析模型对所述历史数据进行分析，得到分析结果；根据所述表情类别标签和所述分析结果匹配对应的音频数据；及获取相匹配的音频数据，将所述音频数据推送至所述用户终端。

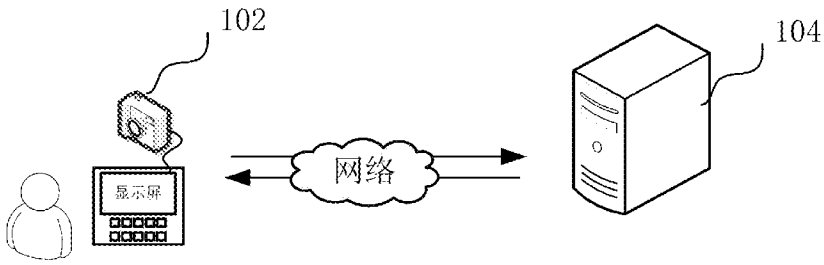
102
104
网络
显示屏

图 1

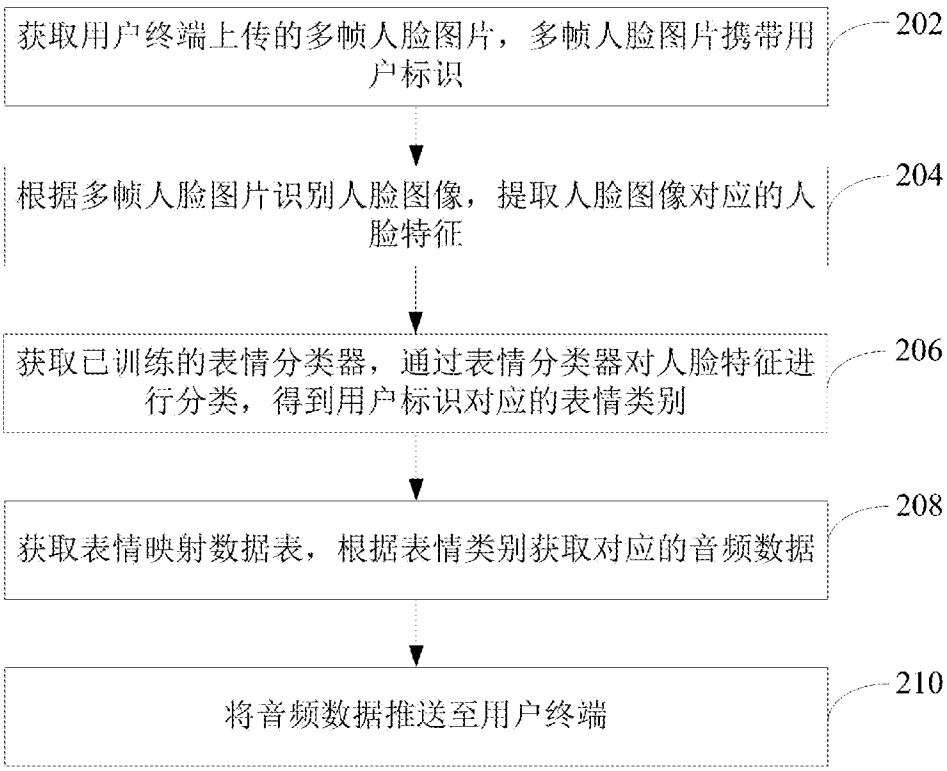
获取用户终端上传的多帧人脸图片，多帧人脸图片携带用户标识
202
根据多帧人脸图片识别人脸图像，提取人脸图像对应的人脸特征
204
获取已训练的表情分类器，通过表情分类器对人脸特征进行分类，得到用户标识对应的表情类别
206
获取表情映射数据表，根据表情类别获取对应的音频数据
208
将音频数据推送至用户终端
210

图 2

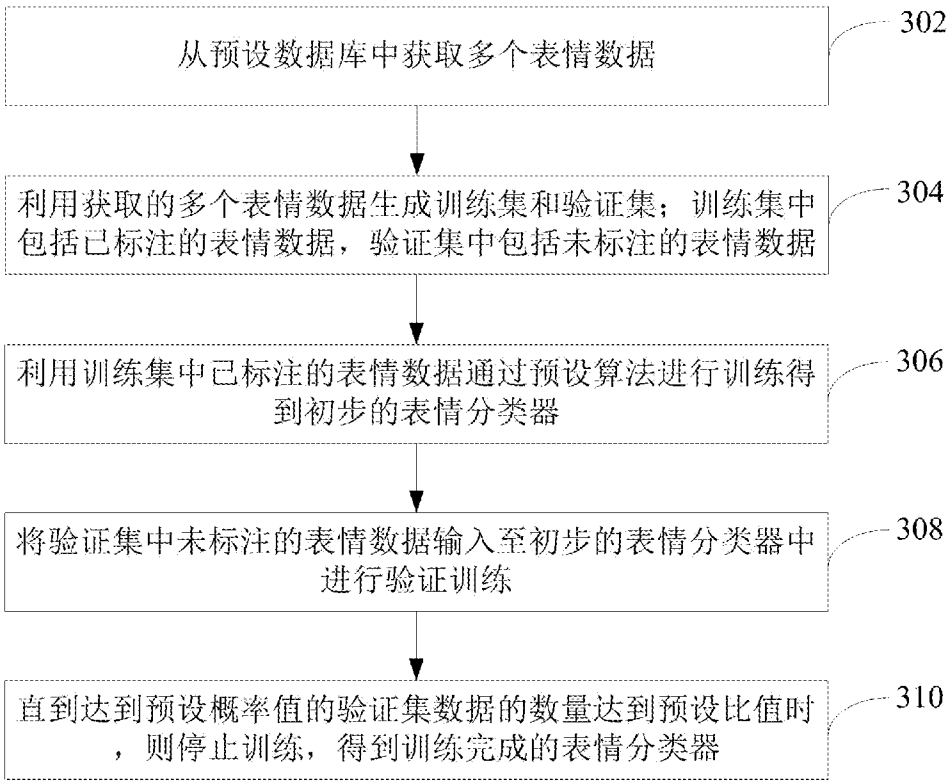
从预设数据库中获取多个表情数据
302
利用获取的多个表情数据生成训练集和验证集；训练集中包括已标注的表情数据，验证集中包括未标注的表情数据
304
利用训练集中已标注的表情数据通过预设算法进行训练得到初步的表情分类器
306
将验证集中未标注的表情数据输入至初步的表情分类器中进行验证训练
308
直到达到预设概率值的验证集数据的数量达到预设比值时，则停止训练，得到训练完成的表情分类器
310

图 3

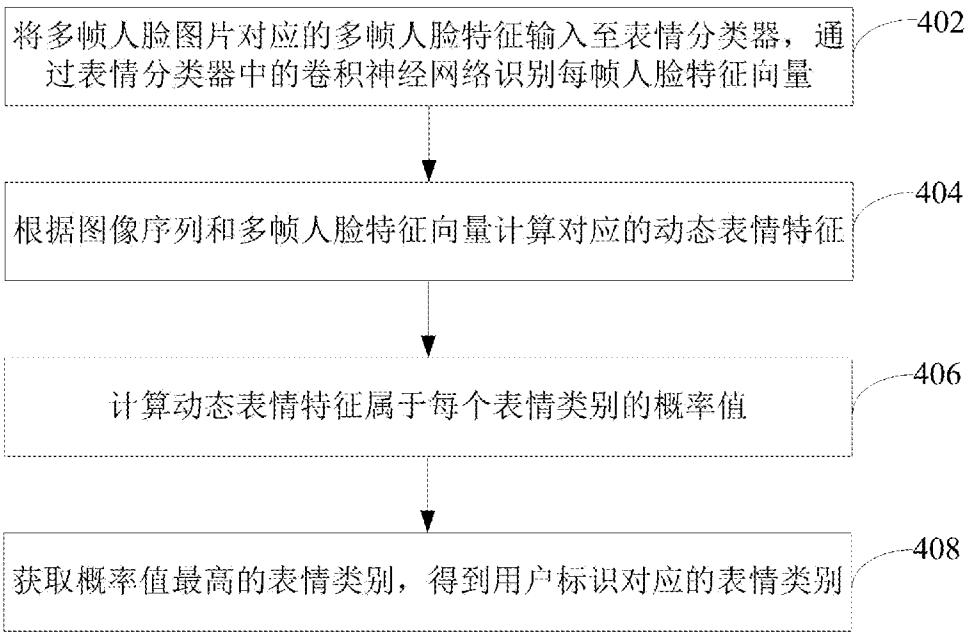
将多帧人脸图片对应的多帧人脸特征输入至表情分类器，通过表情分类器中的卷积神经网络识别每帧人脸特征向量
402
根据图像序列和多帧人脸特征向量计算对应的动态表情特征
404
计算动态表情特征属于每个表情类别的概率值
406
获取概率值最高的表情类别，得到用户标识对应的表情类别
408

图 4

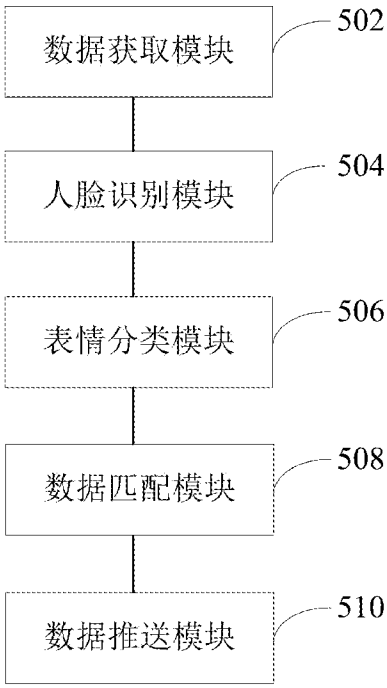
数据获取模块
502
人脸识别模块
504
表情分类模块
506
数据匹配模块
508
数据推送模块
510

图 5

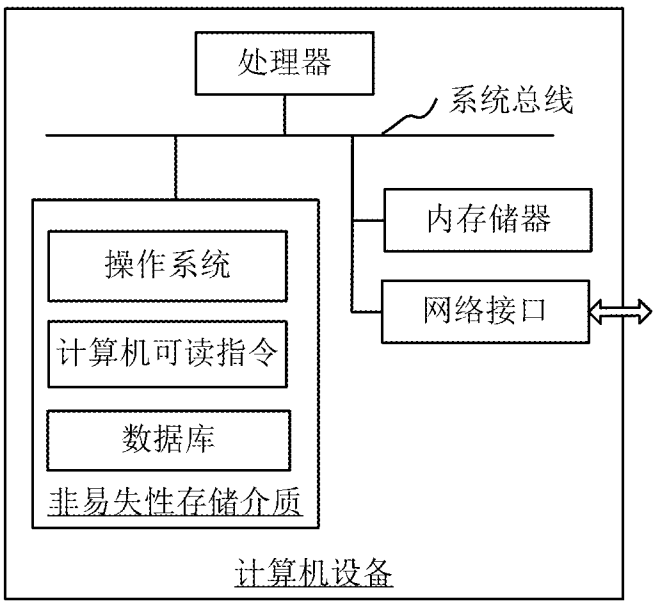
处理器
系统总线
内存储器
操作系统
网络接口
计算机可读指令
数据库
非易失性存储介质
计算机设备

图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/122546

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2006.01)i; G06K 9/62(2006.01)i; G06F 16/63(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K; G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, SIPOABS, CNTXT, DWPI, CNKI: 音频, 音乐, 推送, 推荐, 人脸, 面部, 表情, 分类, 类别, audio, music, push, recommend, face, expression, classify, category

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 108197185 A (NUBIA TECHNOLOGY CO., LTD.) 22 June 2018 (2018-06-22) description, paragraphs 0060-0199	1-20
Y	CN 107862292 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 30 March 2018 (2018-03-30) description, paragraphs 0050-0135	1-20
A	CN 105426404 A (GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CORP., LTD.) 23 March 2016 (2016-03-23) entire document	1-20
PX	CN 109766765 A (ONE CONNECT SMART TECHNOLOGY CO., LTD. (SHENZHEN)) 17 May 2019 (2019-05-17) claims 1-10, and description, paragraphs 0043-0136	1-20

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 February 2020	**06 March 2020**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration** **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/122546

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 108197185 A	22 June 2018	None	
CN 107862292 A	30 March 2018	CN 107862292 B	12 April 2019
		WO 2019095571 A1	23 May 2019
CN 105426404 A	23 March 2016	None	
CN 109766765 A	17 May 2019	None	

国际检索报告

国际申请号
PCT/CN2019/122546

A. 主题的分类

G06K 9/00(2006.01)i; G06K 9/62(2006.01)i; G06F 16/63(2019.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06K; G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用）)

CNABS,SIPOABS,CNTXT,DWPI,CNKI: 音频，音乐，推送，推荐，人脸，面部，表情，分类，类别，audio, music, push, recommend, face, expression, classify, category

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
Y	CN 108197185 A (努比亚技术有限公司) 2018年 6月 22日 (2018-06-22) 说明书第0060-0199段	1-20
Y	CN 107862292 A (平安科技深圳有限公司) 2018年 3月 30日 (2018-03-30) 说明书第0050-0135段	1-20
A	CN 105426404 A (广东欧珀移动通信有限公司) 2016年 3月 23日 (2016-03-23) 全文	1-20
PX	CN 109766765 A (深圳壹账通智能科技有限公司) 2019年 5月 17日 (2019-05-17) 权利要求1-10，说明书第0043-0136段	1-20

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：
“A” 认为不特别相关的表示了现有技术一般状态的文件
“E” 在国际申请日的当天或之后公布的在先申请或专利
“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）
“O” 涉及口头公开、使用、展览或其他方式公开的文件
“P” 公布日先于国际申请日但迟于所要求的优先权日的文件
“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件
“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性
“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性
“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2020年 2月 27日	2020年 3月 6日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 徐薇 电话号码 86-(010)-62411668

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/122546

检索报告引用的专利文件	公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN 108197185 A	2018年 6月 22日	无	
CN 107862292 A	2018年 3月 30日	CN 107862292 B	2019年 4月 12日
		WO 2019095571 A1	2019年 5月 23日
CN 105426404 A	2016年 3月 23日	无	
CN 109766765 A	2019年 5月 17日	无	