

(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
2023 年 12 月 28 日 (28.12.2023)



(10) 国际公布号
WO 2023/246888 A1

(51) 国际专利分类号:

H04N 21/233 (2011.01)

(21) 国际申请号:

PCT/CN2023/101786

(22) 国际申请日:

2023 年 6 月 21 日 (21.06.2023)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

202210722924.7 2022年6月24日 (24.06.2022) CN

(71) 申请人: 中兴通讯股份有限公司 (**ZTE CORPORATION**) [CN/CN]; 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。

(72) 发明人: 惠羿 (**HUI, Yi**); 中国广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦, Guangdong 518057 (CN)。

(74) 代理人: 广州嘉权专利商标事务所有限公司 (**JIAQUAN IP LAW**); 中国广东省广州市天

河区黄埔大道西 100 号富力盈泰广场 A 栋 910, Guangdong 510627 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF,

(54) **Title:** MULTIMEDIA DATA PROCESSING METHOD AND APPARATUS, AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 多媒体数据处理方法、装置和计算机可读存储介质

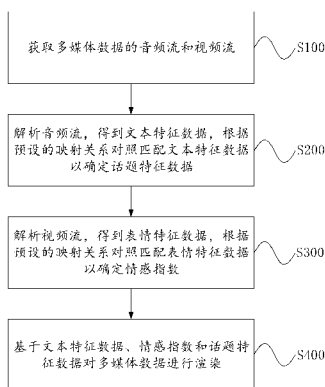


图 1

- S100 Acquire an audio stream and a video stream of multimedia data
- S200 Parse the audio stream to obtain text feature data, and match the text feature data according to a preset mapping relationship, so as to determine topic feature data
- S300 Parse the video stream to obtain expression feature data, and compare the matching expression feature data according to a preset mapping relationship, so as to determine an emotion index
- S400 Render the multimedia data on the basis of the text feature data, the emotion index and the topic feature data

(57) **Abstract:** Disclosed in the present application are a multimedia data processing method and apparatus, and a computer-readable storage medium. The method comprises: acquiring an audio stream and a video stream of multimedia data (S100); parsing the audio stream to obtain text feature data, and matching the text feature data according to a preset mapping relationship, so as to determine topic feature data (S200); parsing the video stream to obtain expression feature data, and comparing the matching expression feature data according to a preset mapping relationship, so as to determine an emotion index (S300); and rendering the multimedia data on the basis of the text feature data, the emotion index and the topic feature data (S400).

(57) 摘要: 本申请公开了多媒体数据处理方法、装置和计算机可读存储介质。该方法包括, 获取多媒体数据的音频流和视频流 (S100); 解析音频流, 得到文本特征数据, 根据预设的映射关系对照匹配文本特征数据以确定话题特征数据 (S200); 解析视频流, 得到表情特征数据, 根据预设的映射关系对照匹配表情特征数据以确定情感指数 (S300); 基于文本特征数据、情感指数和话题特征数据对多媒体数据进行渲染 (S400)。

CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN,
TD, TG)。

根据细则4.17的声明:

— 发明人资格(细则4.17(iv))

本国际公布:

— 包括国际检索报告(条约第21条(3))。

多媒体数据处理方法、装置和计算机可读存储介质

相关申请的交叉引用

本申请基于申请号为 202210722924.7、申请日为 2022 年 6 月 24 日的中国专利申请提出,并要求该中国专利申请的优先权,该中国专利申请的全部内容在此引入本申请作为参考。

技术领域

本申请实施例涉及但不限于数据处理技术领域,特别是涉及一种多媒体数据处理方法、装置和计算机可读取存储介质。

背景技术

随着经济全球化,跨地域的商务沟通越发频繁,目前的远程商务会议多通过视频会议系统进行,通过传输线路及多媒体设备,将声音、影像及文件资料互传,实现即时且互动的沟通,随着增强现实技术的发展以及 5G 技术高通量低延时确定性网络的出现,为视频通话乃至增强现实场景带来了高性能的处理潜力,同时,业务层面也需要商务视频会议能够从数据通道中挖掘提供更多有效的会谈辅助信息,然而,当前主流的商务会议通话产品仅提供视频语音通话控制和媒体通道服务,缺乏对通道数据进行多维度的分析挖掘和二次加工,没有对媒体数据进行有效的挖掘利用。

发明内容

以下是对本文详细描述的主题的概述。本概述并非是为了限制权利要求的保护范围。

本申请实施例提供了一种多媒体数据处理方法、装置和计算机可读存储介质。

第一方面,本申请实施例提供了一种多媒体数据处理方法,包括:获取多媒体数据的音频流和视频流;解析所述音频流,得到文本特征数据,根据预设的映射关系对照匹配所述文本特征数据以确定话题特征数据;解析所述视频流,得到表情特征数据,根据预设的映射关系对照匹配所述表情特征数据以确定情感指数;基于所述文本特征数据、所述情感指数和所述话题特征数据对所述多媒体数据进行渲染。

第二方面,本申请实施例提供了一种多媒体数据处理装置,包括:音频处理模块,被配置为接收并解析音频流,获得文本特征数据;视频处理模块,被配置为接收并解析视频流,得到表情特征数据;映射关系模块,被配置为处理所述文本特征数据得到话题特征数据,处理所述表情特征数据得到情感指数;渲染模块,被配置为将所述文本特征数据、所述话题特征数据和所述情感指数与所述视频流渲染。

第三方面,本申请实施例提供了一种电子设备,包括:存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如上第一方面所述的多媒体数据处理方法。

第四方面,本申请实施例提供了一种计算机可读存储介质,所述计算机可读存储介质存

储有计算机可执行程序，所述计算机可执行程序用于使计算机执行如上第一方面所述的多媒体数据处理方法。

本申请的其它特征和优点将在随后的说明书中阐述，并且，部分地从说明书中变得显而易见，或者通过实施本申请而了解。本申请的目的和其他优点可通过在说明书、权利要求书以及附图中所特别指出的结构来实现和获得。

附图说明

附图用来提供对本申请技术方案的进一步理解，并且构成说明书的一部分，与本申请的实施例一起用于解释本申请的技术方案，并不构成对本申请技术方案的限制。

图 1 是本申请一个实施例提供的一种多媒体数据处理方法的主流程图；

图 2A 是本申请一个实施例提供的一种多媒体数据处理方法的子流程图；

图 2B 是本申请一个实施例提供的解析音频流得到文本特征数据的子流程图；

图 3A 是本申请一个实施例提供的一种多媒体数据处理方法的子流程图；

图 3B 是本申请一个实施例提供的微表情的示意图；

图 3C 是本申请一个实施例提供的解析视频流得到表情特征数据的子流程图；

图 3D 是本申请一个实施例提供的预置微表情组合与情感指数的映射关系的示意图；

图 4A 是本申请一个实施例提供的一种多媒体数据处理方法的子流程图；

图 4B 是本申请一个实施例提供的预置关键短句与场景模板的映射关系的示意图；

图 4C 是本申请一个实施例提供的一种多媒体数据处理方法的子流程图；

图 4D 是本申请一个实施例提供的预置声明语句与命令序列的映射关系的示意图；

图 4E 是本申请一个实施例提供的命令序列类型的示意图；

图 5A 是本申请一个实施例提供的一种多媒体数据处理方法的子流程图；

图 5B 是本申请一个实施例提供的预置敏感语句与渲染集的映射关系的示意图；

图 6 是本申请一个实施例提供的一种多媒体数据处理装置结构的示意图；

图 7 是本申请一个实施例提供的电子设备结构示意图。

具体实施方式

为了使本申请的目的、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，此处所描述的具体实施例仅用以解释本申请，并不用于限定本申请。

应了解，在本申请实施例的描述中，多个（或多项）的含义是两个以上，大于、小于、超过等理解为不包括本数，以上、以下、以内等理解为包括本数。如果有描述到“第一”、“第二”等只是用于区分技术特征为目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量或者隐含指明所指示的技术特征的先后关系。

当今经济全球化，跨地域的商务沟通越发频繁，目前的远程商务会议多通过视频会议系统进行，通过传输线路及多媒体设备，将声音、影像及文件资料互传，实现即时且互动的沟通，随着 XR 技术的发展以及 5G 技术高通量低延时确定性网络的出现，为视频通话乃至 XR 场景带来了高性能的处理潜力，AR、VR、XR 技术的发展使得远程通话技术越来越接近出差现场谈话的效果，虽然 XR 设备的崛起，使得原来一维的语音通话到二维的视频通话，增加了三维

空间通话，但是，媒体数据还没有有效的挖掘利用出来，当前的通话主流产品仅仅提供了视频语音通话控制和媒体通道服务，而对通道数据进行多维度分析的产品少之又少，业务层面也需要商务视频会议能够从数据通道中挖掘提供更多有效的会谈辅助信息。

同时，通讯数据通道和数据处理能力的不断发展，通讯运营商提供能力开放平台，引入第三方开发者来提供丰富多彩的业务的想法由来已久，但是在实践中，国内运营商部署的能力开放平台依然没有取得预定的效果，从根本上来讲，这些开放的能力，还是只能针对原有的语音和视频呼叫做一些简单的控制，从用户体验上来说没有产生突破性的改变。

针对现有视频通话产品的不足之处和业务需要，本申请实施例提供了一种多媒体数据处理方法、装置和计算机可读存储介质，首先，获取多媒体数据的音频流和视频流；解析音频流，得到文本特征数据，根据预设的映射关系对照匹配文本特征数据以确定话题特征数据；解析视频流，得到表情特征数据，根据预设的映射关系对照匹配表情特征数据以确定情感指数；基于文本特征数据、情感指数和话题特征数据对多媒体数据进行渲染。基于此，本申请通过对多人商谈通话的多媒体信息进行解析提取、智能分析和渲染增强，在现有通话产品的语音和视频的基础上，挖掘出多维度的价值信息，包括语音文本、情感指数，并且将语音文本的内容根据预设的映射关系进一步二次加工，使得通话产品能够呈现更丰富智能的信息，同时，通过预设不同的映射关系，能够灵活的处理不同的商务场景和会谈内容，扩展技术方案的应用领域，可以有效的挖掘通话类产品的潜在价值，给通话商谈带来全新的体验，增加商谈业务使用者的会谈竞争力，提高决策效率，赢得更多的话语权，有效抓住商机。

如图1所示，图1是本申请一个实施例提供的一种多媒体数据处理方法的流程图。多媒体数据处理方法包括但不限于如下步骤：

步骤 S100，获取多媒体数据的音频流和视频流；

步骤 S200，解析音频流，得到文本特征数据，根据预设的映射关系对照匹配文本特征数据以确定话题特征数据；

步骤 S300，解析视频流，得到表情特征数据，根据预设的映射关系对照匹配表情特征数据以确定情感指数；

步骤 S400，基于文本特征数据、情感指数和话题特征数据对多媒体数据进行渲染。

可以理解的是，远程商务会议采用视频通话的方式，将会议两端的现场多媒体信息交换传输给对方，网络中传输的多媒体包括音频和视频，在本技术方案中，首先获取多媒体数据的音频流和视频流，将音频流中的声音进行解析提取，获得文本特征数据，例如，音频流中的声音可以是人声，将人声通过语音识别技术转换成文本，即语言文字；类似的，也可以是非人声，根据频谱，识别出其对应的声音类型、名称、特点等，并将声音对应的名称或者动作转换成文本的描述，例如，一种乐器的声音，识别到是钢琴曲，文本特征数据可以包括钢琴曲、钢琴曲的名称、钢琴曲的相关信息等；然后将多媒体数据中的视频流进行解析，得到表情特征数据，表情特征数据包括视频会话中参与人的表情，例如，通过视频流中连续帧的识别，分析出视频中的人脸，并将人脸表情的动态变化提取出来，针对不同的表情，生成不同的表情特征数据，例如，识别出人脸的愤怒表情，通过眉毛，眼睛，脸庞，鼻子，嘴唇等姿态的组合，生成表情特征数据。

可以理解的是，根据预设的映射关系对照匹配文本特征数据以确定话题特征数据，预设的映射关系是将一些文本特征数据跟话题特征数据设置为映射关系，即将一些文本短句跟话

题关键词设置为映射关系,当然,也可以将文本短句跟一些动作命令设置为映射关系,例如,将文本短句匹配为搜索动作,即如果达成匹配,则对匹配到的短句进行检索,将搜索结果作为话题特征数据,当然也可以即将映射关系预设为多级映射关系,即将预置的文本短句跟话题关键词设置为映射关系,进一步将话题关键词跟动作命令设置为映射关系,还可以将动作命令继续跟另一动作命令设置为映射关系,自此,经过多级的映射关系设置,即文本特征数据经过多级映射转化和动作处理后得到话题特征数据,因此,根据预设的映射关系,将文本特征数据中的文本短句跟映射关系中预置的文本短句进行匹配比对,若匹配度达到预设阈值,根据预设的映射关系,将文本特征数据中的文本短句匹配到的预置文本短句所映射的关键词确定为话题特征数据。在一些实施例,文本特征数据可以包含一整篇文本,在进行文本匹配时,将文本特征数据里的文本分散为短句或者短词,分别将分散的短句或者短词跟预置的文本短句进行匹配,预设阈值可以理解为文本特征数据里分散的短句或者短词文本短句有多少比例跟预置的文本短句达成匹配,可以通过设置不同颗粒度的数值,逐步测试合适的预设阈值范围,也可以通过识别的效果判断阈值设置是否合适。根据预设的映射关系对照匹配表情特征数据以确定情感指数,同样的,预设的映射关系是将一些表情特征数据跟情感指数设置为映射关系,即将视频流中提取出来的表情跟情感指数设置为映射关系,情感指数可以根据颜色、数值、图形化等形式表示,例如,将情绪从愤怒到开心设置为0-100分,0分为非常愤怒,100分为非常开心,根据不同的表情,设置对应的情感指数,即可将视频流中提取出来的表情匹配到对应的分数上,同样的,可以通过颜色表示情感指数,不同的颜色对应不同的情绪,通过色彩心理学,因为暖色代表豪放、阳光、热情、热烈、活泼,冷色代表婉约、阴柔、冷静、文雅,因此,可以设置红色代表热烈、轰轰烈烈、冲动;橙色代表快乐、幸福;黄色代表骄傲;绿色代表平和;蓝色代表冷漠、平静、理智、冷酷;紫色代表梦幻;黑色和白色代表恐怖、绝望、悲哀、崇高;灰色代表冷静,将不同的情绪表情根据色彩心理学映射为不同的颜色,根据视频流中提取出来的表情,将之跟预置的表情对照匹配,若匹配度达到预设阈值,根据预设的映射关系,将表情特征数据中的表情匹配到的预置表情所映射的颜色确定为情感指数;但同样,使用图形化也是类似的方式,用图形的圆满度来表示情绪的变化,在此不再做举例。表情特征数据可以是包含一段时间内的表情变化,可能包括多个表情组合,预设阈值可以理解为将此多个表情组合跟预置的表情做比匹配,有多少比例的表情跟预置的表情达成匹配,可以通过设置不同颗粒度的数值,逐步测试合适的预设阈值范围,也可以通过识别的效果反向判断阈值设置是否合适。

可以理解的是,基于文本特征数据、情感指数和话题特征数据对多媒体数据进行渲染。文本特征数据音频流中的语音转化得到的语音文本,情感指数为会谈人的情绪指数,话题特征数据是根据语音文本与预置文本进行比对匹配,根据映射关系,将匹配的预置文本所映射的话题特征数据,或者将匹配的预置文本经过多级映射并执行动作之后得到的话题特征数据,因此,基于文本特征数据、情感指数和话题特征数据对多媒体数据进行渲染,将文本特征数据渲染到视频中,成为会议交谈的字幕,将情感指数渲染到视频中,实时展现参会人员的情绪状态,便于及时根据参会人员的情绪调整话题或者会议内容,将话题特征数据渲染到视频中,是对会议交谈关键信息的二次加工,便于提高会议交谈的效率,或者因为参会人员的知识水平参差不齐或者信息不对称的原因,通过关键信息的二次加工和渲染,及时显示相关话题的相关内容,使参会人员及时把握会议内容,例如,将从音频流得到的文本特征数据,将

文本特征数据经过映射或者多级映射得到的二次加工内容的话题特征数据，和视频流中得到的情感指数，分别跟多媒体数据进行渲染，使得视频流中不只包含会议的画面内容，还包括了多维度的辅助信息，将此视频流返回给媒体面路由，经过传输系统送到会议现场，在会议现场通过 XR 设备，将多媒体数据进行解析和呈现，按照 3Gpp 规范，本技术方案可以部署在路由或媒体处理网元，也可以独立部署，参会人员可以看到原有视频信息，可以听到原有的音频信息，同时，可以看到相关参会人的语音字幕，看到参会人的情感指数，可以即时看到相关话题的总结或者相关信息呈现，当会议在讨论金融安全的话题时，从文本特征数据中分散出若干包含金融相关的短句，根据预设的金融相关短句，例如股市，上市公司名称，上市公司财报，保险公司，证监会等，将文本特征数据中分散出的若干包含金融相关的短句与预设的金融短句进行匹配，当到达预设阈值时，例如一定时间内达到 10 个短句或者关键词的时，即认为达到了预设的阈值，即对匹配成功的短句或者关键词进行映射，映射内容可以是对应的话题关键词，也可以是经过多级动作得到的内容，例如匹配成功的关键词是上市公司名称和财务数据，预置短句映射的是动作，同时动作内容为“检索其近 2 年的收入数据”，因此，基于匹配到的关键词上市公司名称和财务数据，则映射的内容为检索对应上市公司近 2 年的收入数据，在会议现场，立刻会在空间中展现出对应上市公司近 2 年的收入数据，使参会人员更高效的把握话题内容，需要说明的是，可以预设针对各种场景和话题的映射关系，映射关系也可以是多级映射，同时，映射关系可以是关键词，也可以包含执行动作，同样包含多级动作，映射关系经过大量的训练和积累可以不断优化，达到更准确和智能，也可以通过参会人员现场视觉关注内容停留的时间长短，判断映射关系的关联准确度和智能性，实现自动化的优化映射关系和自动录入更新映射关系。

可以理解的是，映射关系包括至少之一：预置微表情组合与情感指数的映射关系；预置关键短句与场景模板的映射关系；预置声明语句与命令序列的映射关系；预置敏感语句与渲染集的映射关系。设置多种维度的映射关系，可以使信息处理时，依照不同维度的信息，采用不同维度的映射关系进行数据处理。

如图 2A 所示，步骤 S200 可以包括但不限于如下子步骤：

步骤 S2001，对音频流进行频率特征学习，得到频率特征频段；

步骤 S2002，利用频率特征频段对音频流过滤，得到若干特征音频流；

步骤 S2003，对特征音频流进行语音识别，得到字幕文本；

步骤 S2004，对特征音频流进行音强分析，得到音强值，当音强值达到预设阈值，则输出字幕附加文本。

可以理解的，如图 2B 所示，对音频流进行频率特征学习，即学习不同人声的频率特征，得到频率特征频段，即不同人声的频率特征所在的频段，通过特征数据，对音频流进行过滤，得到若干特征音频流，即从音频流中将人声分离出来，并且进一步将人声中不同人的声音分离开，得到若干路人声的音频信息，对特征音频流进行语音识别，得到若干人声对应的字幕文本，对音频流进行音强分析，得到音强值。在一些实施例，通过分析人声音频的音调高度或者声音大小，当音强值超过预设阈值时，则输出若干路人声对应的字幕附加文本，即音调或者音量超过预设值的短句或者关键词。

如图 3A 所示，步骤 S300 可以包括但不限于如下子步骤：

步骤 S3001，将视频流输入到预置的深度学习模型，识别获得面部区域坐标；

步骤 S3002, 根据面部区域坐标, 对视频流中面部区域的连续帧进行微表情识别分块, 得到微表情组合。

如图 3B 所示, 面部表情涉及多个部位的动作, 其中“微表情”自然对应着的是“宏表情”。宏表情在我们平时比较容易观察到的, 面部表情持续时间在 0.5-4 秒范围之内, 且参与面部表情动作的面部肌肉群收缩或舒张幅度较大。微表情不但受思维意识控制, 而且持续的时间非常短, 人的意识还没有来得及去控制它, 就已经被暴露出来, 所以它一定程度上反应了真实的情绪。正是由于微表情的这个特点, 它在刑侦、安全、司法、谈判等领域都有非常重要的应用。微表情是一种特殊的面部表情, 与普通的表情相比, 微表情主要有以下特点:

1. 持续时间短, 通常只有 $1/25s \sim 1/3s$;
2. 动作强度低, 难以察觉;
3. 在无意识状态下产生, 通常难以掩饰或伪装;
4. 对微表情的分析通常需要在视频中, 而普通表情在图像中就可以分析。

微表情的识别, 它最开始是通过人工训练的, 大概训练一个半小时, 可以提高到 30%-40% 的准确性, 但心理学家论证了人工识别最高不会超过 47%。后来随着心理学实验逐渐演进到计算机应用上面, 微表情有了 AU (Action Units) 组合, 例如, 如果是高兴, 宏表情的高兴是 AU6+AU12, 微表情是 AU6 或者是 AU12, 或者是 AU6+AU12。第二是微表情运动, 它是局部的肌肉块运动, 而不是两个肌肉块同时运动, 比如高兴, 运动的是 AU6 或者是 AU12。但如果强度比较大, 可能会出现 AU6 和 AU12 同时发生的情况。另外宏表情和微表情有一些联系, 就是微表情的 AU 可能是宏表情 AU 的子集, 因此, 通过微表情识别技术 MER (Micro expression recognition) 可以从视频中提取出微表情。通过微表情组合, 即可准确描述参会人员的表情变化, 侧面反应出参会人员的情绪变化, 可以进一步分析出参会人员讨论话题的兴趣程度。因此, 如图 3C 所示, 将视频流输入到预置的深度学习模型, 在一些实施例, 可以采用卷积神经网络 (Convolutional Neural Networks, CNN) 模型, 先进行学习训练, 得到成熟的 CNN 模型, 将视频流输入到 CNN 模型, 进行人脸识别, 识别获得面部区域坐标, 获得视频中各个会谈人员面部的区域坐标, 使用 MER 分块器, 对视频人脸区域的连续帧进行 MER 分块, 进一步 MER 序列解析, 对连续时序图片的变化进行输出, 得到微表情 AU 的组合序列, 即微表情组合。

可以理解的是, 如图 3D 所示, 预置微表情组合与情感指数的映射关系, 即微表情组合由若干微表情组成, 例如, 微表情 1、微表情 2 和微表情 3, 其中, 微表情 1 为 AU6, 微表情 2 为 AU7, 微表情 3 为 AU12, 即微表情组合为 AU6+AU7+AU12, 同样的, 可以对微表情进行排列组合, 得到微表情组合的集合, 根据微表情集合所对应表达的情绪, 映射为对应的情感指数, 因此, 通过对视频流中的参会人员的表情提取, 得到微表情组合, 并根据预置为表情组合的映射关系, 得到情感指数, 情感指数可以根据颜色、数值、图形化等形式表示, 例如, 将情绪从愤怒到开心设置为 0-100 分, 0 分为非常愤怒, 100 分为非常开心, 根据不同的表情, 设置对应的情感指数, 即可将视频流中提取出来的表情匹配到对应的分数上, 同样的, 可以通过颜色表示情感指数, 不同的颜色对应不同的情绪, 通过色彩心理学, 因为暖色代表豪放、阳光、热情、热烈、活泼, 冷色代表婉约、阴柔、冷静、文雅, 因此, 可以设置红色代表热烈、轰轰烈烈、冲动; 橙色代表快乐、幸福; 黄色代表骄傲; 绿色代表平和; 蓝色代表冷漠、平静、理智、冷酷; 紫色代表梦幻; 黑色和白色代表恐怖、绝望、悲哀、崇高; 灰色代表冷

静，将不同的微表情组合根据色彩心理学映射为不同的颜色，即可将微表情组合映射为情感指数。

如图 4A 所示，步骤 S200 还包括但不限于如下子步骤：

步骤 S2005，将字幕文本通过自然语言处理，获得短句序列；

步骤 S2006，利用映射关系，将短句序列与预置关键短句进行对照匹配，得到对应的场景模板；

步骤 S2007，依据对应的场景模板，对字幕文本、字幕附加文本进行深层处理获得话题特征数据。

可以理解的是，将字幕文本进行语义处理，在一些实施例中，通过自然语言处理 (Natural language processing)，将字幕文本按照语义分割成若干短句序列，利用映射关系，将短句序列与预置关键短句进行对照匹配，当匹配度达到预设阈值时，得到对应的场景模板。

如图 4B 所示，预置关键短句与场景模板的映射关系，至少一个预置关键短句与一个场景映射，例如，预置一个金融领域的话题场景，预置的关键短句为上证指数，二级市场，上海证券交易所，深圳证券交易所，两市资金流入流出，A 股，6 个预置关键短句与金融场景映射，将视频会议音频流提取出来的字幕文本按照语义分割成若干短句序列，将此若干短句序列跟 6 个预置关键短句进行比对匹配，计算相似度，当相似度达到预设阈值时，判断认为当前会议话题是金融话题，则进入金融场景。依据对应的场景模板，对字幕文本、字幕附加文本进行深层处理获得话题特征数据。在一些实施例中，在金融场景下，根据预置的映射关系，进一步将字幕文本和字幕附加文本进行语义分割，与预置的语句进行比对，得到对数据加工的动作指令，通过映射的动作，将字幕文本和字幕附加文本加工进行二次加工，将二次加工结果输出渲染。需要说明的是，阈值可以根据场景匹配的严格程度和预置关键短句的数量来设置，例如，同一个场景下，预置关键短句数量越多，要求达到的匹配度越高，则阈值越高，越难以与场景匹配，但是准确度会越高，因此，可以通过大量的数据训练，找到合适的阈值范围，兼顾准确度和实用性。需要说明的是，本方案中预置多种场景，多个预置关键短句映射一个场景，可以设置不同维度，不同级别的场景，来应用于不同类型的会议主题。

如图 4C 所示，步骤 S2007 还包括但不限于如下子步骤：

步骤 S20071，利用映射关系，将短句序列和字幕附加文本与预置声明语句进行对照匹配，得到对应的命令序列；

步骤 S20072 执行命令序列，获得话题特征数据。

可以理解的是，利用映射关系，即预置声明语句与命令序列的映射关系，将短句序列和字幕附加文本与预置声明语句进行对照匹配，得到对应的命令序列。

如图 4D 所示，预置声明语句与命令序列的映射关系，至少一个预置声明语句与一个命令序列映射，命令序列包含一个动作组，即若干数据处理动作，例如，预置一个 5G 通信领域的话题场景，预置的关键短句为 5G 组网，非独立组网 NSA，独立组网 SA，控制承载分离，移动边缘计算，网络切片，6 个预置关键短句与 5G 通信场景映射，将视频会议音频流提取出来的字幕文本按照语义分割成若干短句序列，将此若干短句序列跟 6 个预置关键短句进行比对匹配，计算相似度，当相似度达到预设阈值时，判断认为当前会议话题是 5G 通信话题，则进入 5G 通信场景，在 5G 通信场景内，对字幕文本、字幕附加文本做进一步的匹配映射。在一些实施例中，利用预置声明语句与命令序列的映射关系，将短句序列和字幕附加文本与预置声明

语句进行对照匹配,得到对应的命令序列,例如,预置声明语句为5G专利数量,5G标准必要专利,5G专利持有量,5G技术公司,5G市场规模,5个预置声明语句,同时,5个预置声明语句所映射的命令序列可以设置为:

动作1:搜索引擎检索预置声明语句;

动作2:提取检索结果前五条网址的内容简介或者关键词;

动作3:在专利数据库检索网页中提取的关键词;

动作4:将检索结果去重后输出。

命令序列包含4个动作,通过此映射关系,将若干短句序列与6个5G通信场景下的5个预置声明语句进行比对匹配,当相似度达到预设值时,即认为匹配成功,因此执行预置声明语句所映射的命令序列,即4个动作,最后输出动作结果。需要说明的是,阈值可以根据匹配的严格程度和预置声明语句的数量来设置,例如,同一个命令序列下,预置声明语句数量越多,要求达到的匹配度越高,则阈值越高,越难以与命令序列匹配,但是准确度会越高,因此,可以通过大量的数据训练,找到合适的阈值范围,兼顾准确度和实用性。需要说明的是,本方案中预置多种命令序列,多个命令序列映射一个命令序列,可以设置不同维度,不同级别的命令序列,来应用于不同类型的数据二级加工场景。

如图4E所示,步骤S20071还包括但不限于如下至少一种动作:

当命令序列包含在线检索指令,则检索在线数据库,得到在线检索数据集;

当命令序列包含离线检索指令,则检索离线数据库,得到离线检索数据集;

当命令序列包含数据加工指令,则对数据进行二次加工,得到二次加工数据。

在线检索数据库,可以采用谷歌搜索引擎、百度搜索引擎、必应搜索引擎、搜狗搜索引擎、好搜搜索引擎、神马搜索引擎、雅虎搜索引擎中的一种或多种结合,当然也可以使用细分专业领域的搜索引擎,例如,专利检索领域的专用数据库,文献检索领域的专用数据库等;离线数据库是本地数据库,可以为局域网内的内部数据库,针对本技术方案所使用的领域或用户自有数据库,可以灵活设置;当命令序列包含数据加工指令,则对数据进行二次加工,二次加工的方法依然可以采用映射关系,进行多重映射,对数据进行多次多维度的处理。

如图5A所示,步骤S400还包括但不限于如下子步骤:

步骤S4001,利用字幕文本,获得特征音频流启动时对应的声音时间序列;

步骤S4002,利用微表情组合,获得嘴型表情启动时对应的嘴型时间序列;

步骤S4003,判断声音时间序列和嘴型时间序列的一致性,将字幕文本、情感指数和面部区域坐标在时序上对齐;

步骤S4004,利用映射关系,将短句序列和预置敏感语句进行对照匹配,得到对应的渲染集;

步骤S4005,根据渲染集的配置,将字幕文本、情感指数和话题特征数据与视频流渲染。

可以理解的是,利用字幕文本,获得特征音频流启动时对应的声音时间序列,字幕文本提取时带有时间标记,可以确定文本所对应的时间;利用微表情组合,获得嘴型表情启动时对应的嘴型时间序列,从视频流中提取微表情时,根据时间序列顺序提取,因此微表情组合也带有时间标记,同时,因为视频流和音频流是同一多媒体源,具有一致的时序,因此,在数据处理时,由于音频流和视频流分路处理,并且经过二次数据处理,可能会出现数据时序

不同步,因此,判断声音时间序列和嘴型时间序列的一致性,将字幕文本、情感指数和面部区域坐标在时序上对齐,字幕文本对应的是声音时间序列,情感指数对应的是嘴型时间序列,面部区域坐标对应的是嘴型时间序列,当判断声音时间序列和嘴型时间序列一致时,则字幕文本、情感指数和面部区域坐标在时序上也是对齐的。

如图 5B 所示,利用映射关系,将短句序列和预置敏感语句进行对照匹配,得到对应的渲染集,预置敏感语句与渲染集设置有映射关系,至少一句预置敏感语句与渲染集映射,根据使用场景的需要,可以预设不同维度的预置敏感语句作为匹配参照。

渲染集的配置包括渲染坐标、渲染时间、渲染颜色主题至少之一,不同的渲染集,在将字幕文本、情感指数和话题特征数据与视频流渲染时,会呈现出不同的视觉效果。例如,在 XR 会议中,根据渲染集的不同配置,字幕文本、情感指数和话题特征数据与视频流渲染后,会议参与人会在不同的人脸庞附近看到不同颜色的字幕文本,不同风格的情感指数,不同展示时间的话题特征数据。参会人员 XR 设备中可以看到不同会谈人的附近区域,有显示板或者弹幕弹出各自的会谈字幕,看到参与人的情感指数的显示,会谈期间的关键信息经过二次加工后会反馈到视觉空间中同时,不同级别的关键信息显示时长、字体颜色、显示位置均表现不同,呈现的方式可以是图标或者数字,也可以是其他设计,同时,可以通过管理界面,对不同的关键词句进行分组,分组后的集合可以设置响应的处理加工动作,可以进行不同的渲染效果设置。

如图 6 所示,本申请实施例还提供了一种多媒体数据处理装置,可以通过对多人商谈通话的多媒体信息进行解析提取、智能分析和渲染增强,在现有通话产品的语音和视频的基础上,挖掘出多维度的价值信息,包括语音文本、情感指数,并且将语音文本的内容根据预设的映射关系进一步二次加工,使得通话产品能够呈现更丰富智能的信息,同时,通过预设不同的映射关系,能够灵活的处理不同的商务场景和会谈内容,扩展技术方案的应用领域,可以有效的挖掘通话类产品的潜在价值,给通话商谈带来全新的体验,增加商谈业务使用者的会谈竞争力,提高决策效率,赢得更多的话语权,有效抓住商机。

该多媒体数据处理装置包括:

音频处理模块 501,被配置为接收音频数据,将音频流中的声音进行解析提取,获得文本特征数据;

视频处理模块 502,被配置为接收视频数据,对视频流进行解析,得到表情特征数据,表情特征数据包括视频会话中参与人的表情,通过视频流中连续帧的识别,分析出视频中的人脸,并将人脸表情的动态变化提取出来,针对不同的表情,生成不同的表情特征数据;

映射关系模块 503,根据预设的映射关系,将文本特征数据转换为话题特征数据;将表情特征数据转换为情感指数;

渲染模块 504,将文本特征数据、情感指数和话题特征数据渲染输出。

在一实施方式中,音频处理模块接收音频数据,将音频流中的声音进行解析提取,获得文本特征数据;视频处理模块接收视频数据,对视频流进行解析,得到表情特征数据,表情特征数据包括视频会话中参与人的表情,通过视频流中连续帧的识别,分析出视频中的人脸,并将人脸表情的动态变化提取出来,针对不同的表情,生成不同的表情特征数据;映射关系模块根据预设的映射关系,将文本特征数据转换为话题特征数据,将表情特征数据转换为情感指数;渲染模块将文本特征数据、情感指数和话题特征数据渲染输出。基于此,本申请通过

对多人商谈通话的多媒体信息进行解析提取、智能分析和渲染增强,在现有通话产品的语音和视频的基础上,挖掘出多维度的价值信息,包括语音文本、情感指数,并且将语音文本的内容根据预设的映射关系进一步二次加工,使得通话产品能够呈现更丰富智能的信息,同时,通过预设不同的映射关系,能够灵活的处理不同的商务场景和会谈内容,扩展技术方案的应用领域,可以有效的挖掘通话类产品的潜在价值,给通话商谈带来全新的体验,增加商谈业务使用者的会谈竞争力,提高决策效率,赢得更多的话语权,有效抓住商机。

如图7所示,本申请实施例还提供了一种电子设备。可以通过对多人商谈通话的多媒体信息进行解析提取、智能分析和渲染增强,在现有通话产品的语音和视频的基础上,挖掘出多维度的价值信息,包括语音文本、情感指数,并且将语音文本的内容根据预设的映射关系进一步二次加工,使得通话产品能够呈现更丰富智能的信息,同时,通过预设不同的映射关系,能够灵活的处理不同的商务场景和会谈内容,扩展技术方案的应用领域,可以有效的挖掘通话类产品的潜在价值,给通话商谈带来全新的体验,增加商谈业务使用者的会谈竞争力,提高决策效率,赢得更多的话语权,有效抓住商机。

该电子设备包括:

处理器,可以采用通用的CPU(Central Processing Unit,中央处理器)、微处理器、应用专用集成电路(Application Specific Integrated Circuit,ASIC)、或者一个或多个集成电路等方式实现,被配置为执行相关程序,以实现本申请实施例所提供的技术方案;

存储器,可以采用只读存储器(Read Only Memory,ROM)、静态存储设备、动态存储设备或者随机存取存储器(Random Access Memory,RAM)等形式实现。存储器可以存储操作系统和其他应用程序,在通过软件或者固件来实现本说明书实施例所提供的技术方案时,相关的程序代码保存在存储器中,并由处理器来调用执行本申请实施例的目标数据的排序方法;

输入/输出接口,被配置为实现信息输入及输出;

通信接口,被配置为实现本设备与其他设备的通信交互,可以通过有线方式(例如USB、网线等)实现通信,也可以通过无线方式(例如移动网络、WIFI、蓝牙等)实现通信;

总线,在设备的各个组件(例如处理器、存储器、输入/输出接口和通信接口)之间传输信息;

其中处理器、存储器、输入/输出接口和通信接口通过总线实现彼此之间在设备内部的通信连接。

在一实施方式中,该电子设备包括:一个或多个处理器和存储器,图7中以一个处理器及存储器为例。处理器和存储器可以通过总线或者其他方式连接,图7中以通过总线连接为例。

存储器作为一种非暂态计算机可读存储介质,可用于存储非暂态软件程序以及非暂态性计算机可执行程序,如上述本申请实施例中的多媒体数据处理方法。处理器通过运行存储在存储器中的非暂态软件程序以及程序,从而实现上述本申请实施例中的多媒体数据处理方法。

存储器可以包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需要的应用程序;存储数据区可存储执行上述本申请实施例中的多媒体数据处理方法所需的数据等。此外,存储器可以包括高速随机存取存储器,还可以包括非暂态存储器,例如至少一个磁盘存储器件、闪存器件、或其他非暂态固态存储器件。在一些实施方式中,存储器包括相对于处理器远程设置的存储器,这些远程存储器可以通过网络连接至该多媒体

数据处理装置。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

实现上述本申请实施例中的多媒体数据处理方法所需的非暂态软件程序以及程序存储在存储器中，当被一个或者多个处理器执行时，执行上述本申请实施例中的多媒体数据处理方法，例如，执行以上描述的图 1 中的方法步骤 S100 至步骤 S400，图 2A 中的方法步骤 S2001 至步骤 S2004，图 3A 中的方法步骤 S3001 至步骤 S3002，图 4A 中的方法步骤 S2005 至步骤 S2007。获取多媒体数据的音频流和视频流；解析音频流，得到文本特征数据，根据预设的映射关系对照匹配文本特征数据以确定话题特征数据；解析视频流，得到表情特征数据，根据预设的映射关系对照匹配表情特征数据以确定情感指数；基于文本特征数据、情感指数和话题特征数据对多媒体数据进行渲染。基于此，本申请通过对多人商谈通话的多媒体信息进行解析提取、智能分析和渲染增强，在现有通话产品的语音和视频的基础上，挖掘出多维度的价值信息，包括语音文本、情感指数，并且将语音文本的内容根据预设的映射关系进一步二次加工，使得通话产品能够呈现更丰富智能的信息，同时，通过预设不同的映射关系，能够灵活的处理不同的商务场景和会谈内容，扩展技术方案的应用领域，可以有效的挖掘通话类产品的潜在价值，给通话商谈带来全新的体验，增加商谈业务使用者的会谈竞争力，提高决策效率，赢得更多的话语权，有效抓住商机。

此外，本申请实施例还提供了一种计算机可读存储介质，该计算机可读存储介质存储有计算机可执行程序，该计算机可执行程序被一个或多个控制处理器执行，例如，被图 7 中的一个处理器执行，可使得上述一个或多个处理器执行上述本申请实施例中的多媒体数据处理方法，例如，执行以上描述的图 1 中的方法步骤 S100 至步骤 S400，图 2A 中的方法步骤 S2001 至步骤 S2004，图 3A 中的方法步骤 S3001 至步骤 S3002，图 4A 中的方法步骤 S2005 至步骤 S2007。获取多媒体数据的音频流和视频流；解析音频流，得到文本特征数据，根据预设的映射关系对照匹配文本特征数据以确定话题特征数据；解析视频流，得到表情特征数据，根据预设的映射关系对照匹配表情特征数据以确定情感指数；基于文本特征数据、情感指数和话题特征数据对多媒体数据进行渲染。基于此，本申请通过对多人商谈通话的多媒体信息进行解析提取、智能分析和渲染增强，在现有通话产品的语音和视频的基础上，挖掘出多维度的价值信息，包括语音文本、情感指数，并且将语音文本的内容根据预设的映射关系进一步二次加工，使得通话产品能够呈现更丰富智能的信息，同时，通过预设不同的映射关系，能够灵活的处理不同的商务场景和会谈内容，扩展技术方案的应用领域，可以有效的挖掘通话类产品的潜在价值，给通话商谈带来全新的体验，增加商谈业务使用者的会谈竞争力，提高决策效率，赢得更多的话语权，有效抓住商机。

本领域普通技术人员可以理解，上文中所公开方法中的全部或某些步骤、系统可以被实施为软件、固件、硬件及其适当的组合。某些物理组件或所有物理组件可以被实施为由处理器，如中央处理器、数字信号处理器或微处理器执行的软件，或者被实施为硬件，或者被实施为集成电路，如专用集成电路。这样的软件可以分布在计算机可读介质上，计算机可读介质可以包括计算机存储介质（或非暂时性介质）和通信介质（或暂时性介质）。如本领域普通技术人员公知的，术语计算机存储介质包括在用于存储信息（诸如计算机可读程序、数据结构、程序模块或其他数据）的任何方法或技术中实施的易失性和非易失性、可移除和不可移除介质。计算机存储介质包括但不限于 RAM、ROM、EEPROM、闪存或其他存储器技术、CD-ROM、

数字多功能盘（DVD）或其他光盘存储、磁盒、磁带、磁盘存储或其他磁存储装置、或者可以用于存储期望的信息并且可以被计算机访问的任何其他的介质。此外，本领域普通技术人员公知的是，通信介质通常包含计算机可读程序、数据结构、程序模块或者诸如载波或其他传输机制之类的调制数据信号中的其他数据，并且可包括任何信息递送介质。

以上是对本申请的部分实施进行说明，但本申请并不局限于上述实施方式，熟悉本领域的技术人员在不违背本申请本质的共享条件下还可作出种种等同的变形或替换，这些等同的变形或替换均包括在本申请权利要求所限定的范围内。

权 利 要 求 书

1. 一种多媒体数据处理方法，包括：

获取多媒体数据的音频流和视频流；

解析所述音频流，得到文本特征数据，根据预设的映射关系对照匹配所述文本特征数据以确定话题特征数据；

解析所述视频流，得到表情特征数据，根据预设的映射关系对照匹配所述表情特征数据以确定情感指数；

基于所述文本特征数据、所述情感指数和所述话题特征数据对所述多媒体数据进行渲染。

2. 根据权利要求1所述的方法，其中，所述映射关系包括如下至少之一：

预置微表情组合与情感指数的映射关系；或

预置关键短句与场景模板的映射关系；或

预置声明语句与命令序列的映射关系；或

预置敏感语句与渲染集的映射关系。

3. 根据权利要求2所述的方法，其中，所述解析所述音频流，得到文本特征数据，包括：

对所述音频流进行频率特征学习，得到频率特征频段；

利用所述频率特征频段对所述音频流过滤，得到若干特征音频流；

对所述特征音频流进行语音识别，得到字幕文本；

对所述特征音频流进行音强分析，得到音强值，当所述音强值达到预设阈值，则输出字幕附加文本。

4. 根据权利要求3所述的方法，其中，所述解析所述视频流，得到表情特征数据，包括：

将所述视频流输入到预置的深度学习模型，识别获得面部区域坐标；

根据所述面部区域坐标，对所述视频流中面部区域的连续帧进行微表情识别分块，得到微表情组合。

5. 根据权利要求4所述的方法，其中，所述根据预设的映射关系对照匹配所述表情特征数据以确定情感指数，包括：

利用所述映射关系，将所述微表情组合与所述预置微表情组合进行对照匹配，得到对应的情感指数。

6. 根据权利要求3所述的方法，其中，所述根据预设的映射关系对照匹配所述文本特征数据以确定话题特征数据，包括：

将所述字幕文本通过自然语言处理，获得短句序列；

利用所述映射关系，将所述短句序列与所述预置关键短句进行对照匹配，得到对应的所述场景模板；

依据对应的所述场景模板，对所述字幕文本、所述字幕附加文本进行深层处理获得话题特征数据。

7. 根据权利要求6所述的方法，其中，所述依据对应的所述场景模板，对所述字幕文本、所述字幕附加文本进行深层处理获得话题特征数据，包括：

利用所述映射关系，将所述短句序列和所述字幕附加文本与所述预置声明语句进行对照匹配，得到对应的所述命令序列；

执行所述命令序列，获得话题特征数据。

8. 根据权利要求 7 所述的方法，其中，所述执行所述命令序列，获得话题特征数据，包括如下至少之一：

当所述命令序列包含在线检索指令，则检索在线数据库，得到在线检索数据集；或
当所述命令序列包含离线检索指令，则检索离线数据库，得到离线检索数据集；或
当所述命令序列包含数据加工指令，则对数据进行二次加工，得到二次加工数据。

9. 根据权利要求 4 至 8 任一所述的方法，其中，所述基于所述文本特征数据、所述情感指数和所述话题特征数据对所述多媒体数据进行渲染，包括：

利用所述字幕文本，获得所述特征音频流启动时对应的声音时间序列；

利用所述微表情组合，获得嘴型表情启动时对应的嘴型时间序列；

判断所述声音时间序列和所述嘴型时间序列的一致性，将所述字幕文本、所述情感指数和所述面部区域坐标在时序上对齐；

利用所述映射关系，将所述短句序列和所述预置敏感语句进行对照匹配，得到对应的所述渲染集；

根据所述渲染集的配置，将所述字幕文本、所述情感指数和话题特征数据与所述视频流渲染。

10. 根据权利要求 9 所述的方法，其中，所述渲染集的配置，包括如下至少之一：

渲染坐标；或渲染时间；或渲染颜色主题。

11. 一种多媒体数据处理装置，包括：

音频处理模块，被配置为接收并解析音频流，获得文本特征数据；

视频处理模块，被配置为接收并解析视频流，得到表情特征数据；

映射关系模块，被配置为处理所述文本特征数据得到话题特征数据，处理所述表情特征数据得到情感指数；

渲染模块，被配置为将所述文本特征数据、所述话题特征数据和所述情感指数与所述视频流渲染。

12. 一种电子设备，包括：存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现如权利要求 1 至 10 中任意一项所述的多媒体数据处理方法。

13. 一种计算机可读存储介质，存储有计算机可执行程序，所述计算机可执行程序用于使计算机执行如权利要求 1 至 10 任意一项所述的多媒体数据处理方法。

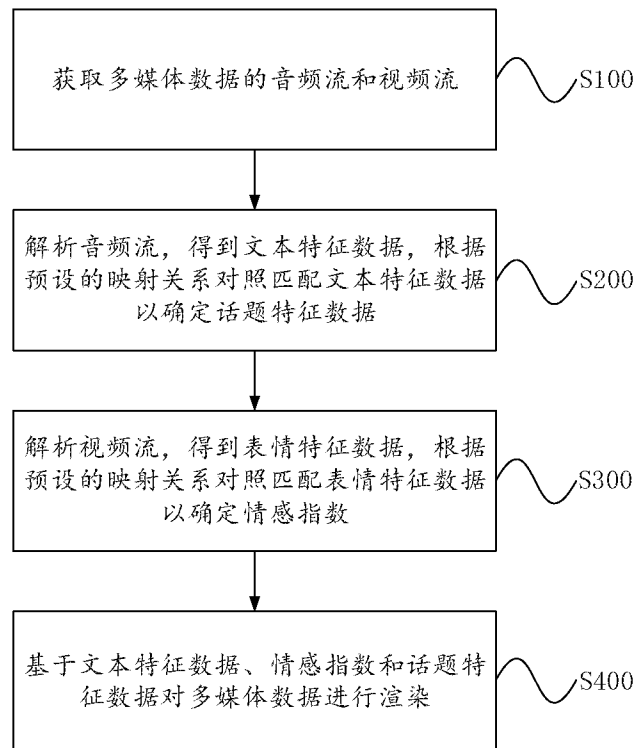


图 1

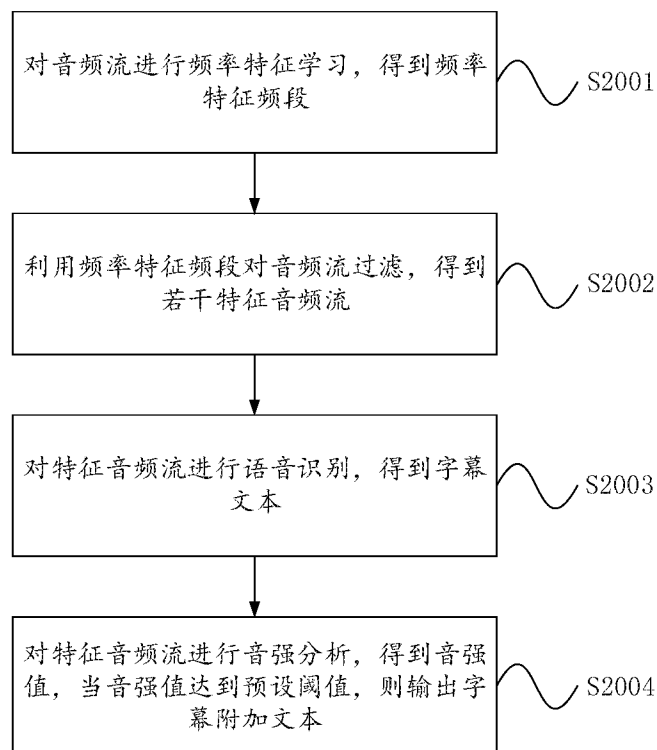


图 2A

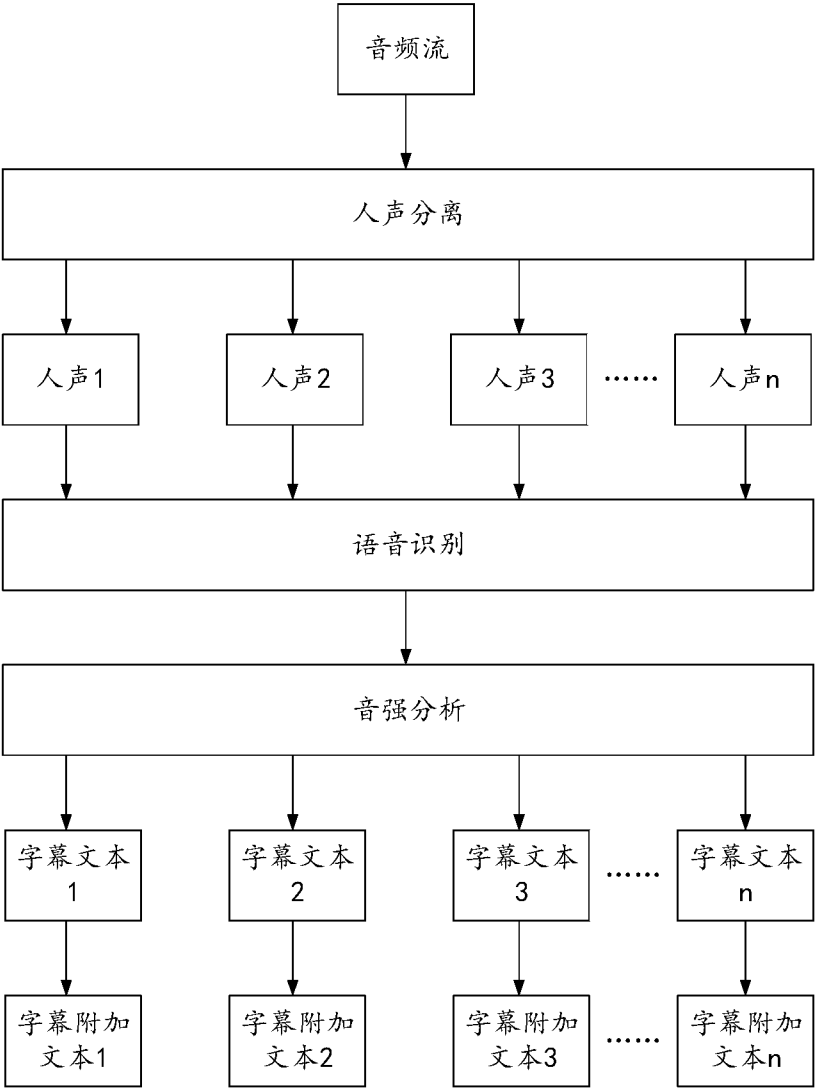


图 2B

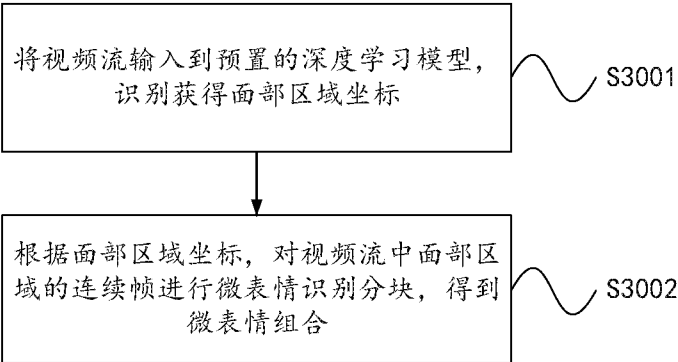


图 3A

面部表情

AU	描述	面部表情	AU	描述	面部表情	AU	描述	面部表情
1	眉毛内侧上升		2	外侧眉毛上升		4	眉毛压低	
5	眉毛压低		6	脸颊上升		7	脸颊变紧	
9	鼻子皱起		10	上嘴唇上升		11	鼻翼肉加深	
12	嘴角上拉		13	面颊拉伸		14	出现酒窝	
15	嘴角压低		16	下嘴唇压低		17	下巴上升	
18	嘴唇鼓起		20	嘴唇伸展		22	嘴唇呈漏斗型	
23	嘴唇变紧		24	嘴唇收缩		25	嘴唇分开	
26	下巴降低		27	嘴巴张开		28	嘴唇吸紧	
41	脸颊下垂		42	眯眼		44	斜视	

图 3B

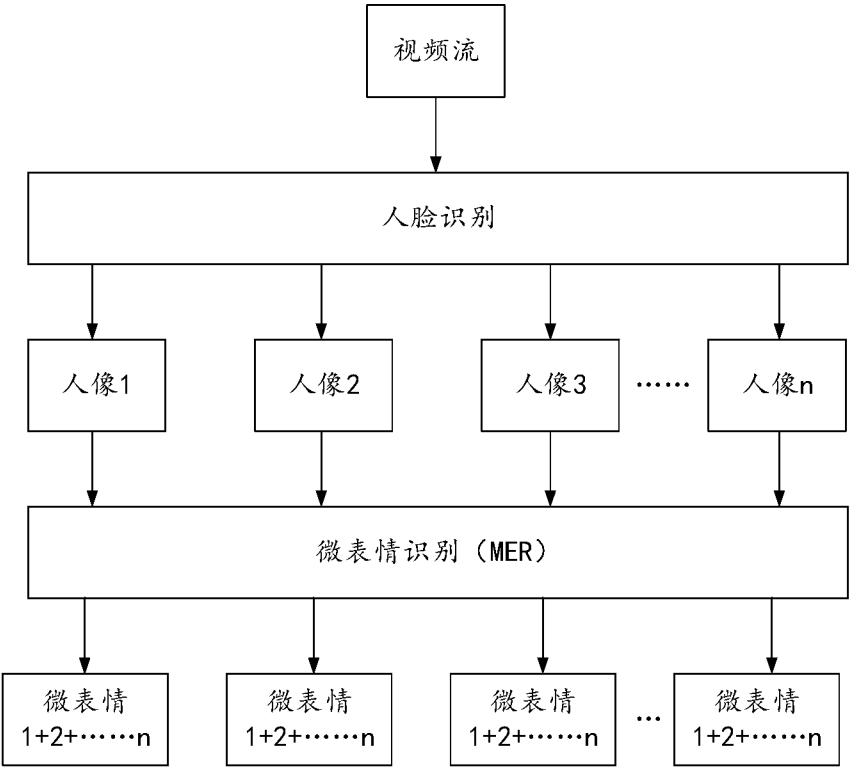


图 3C

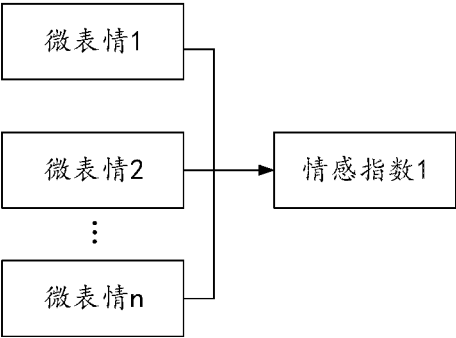


图 3D

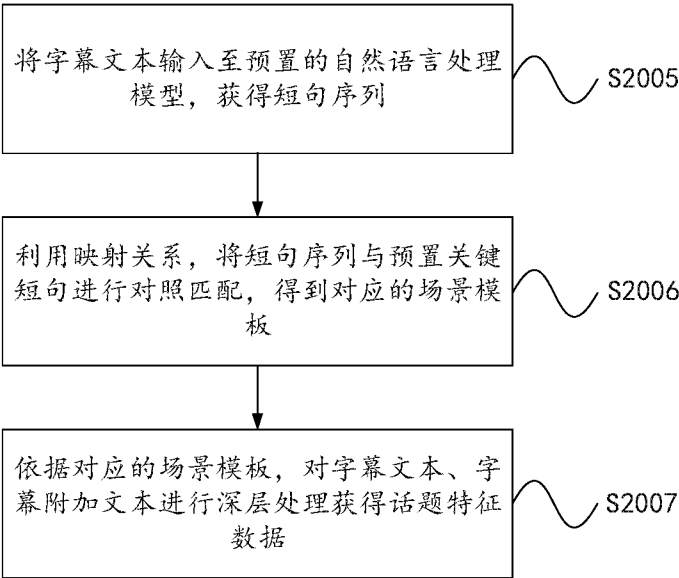


图 4A

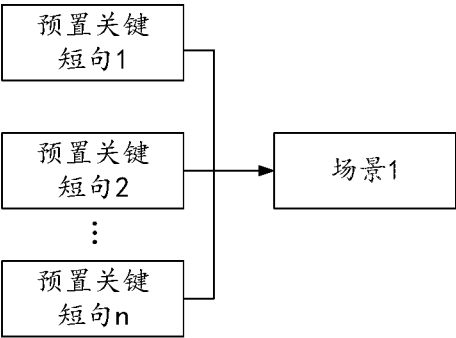


图 4B

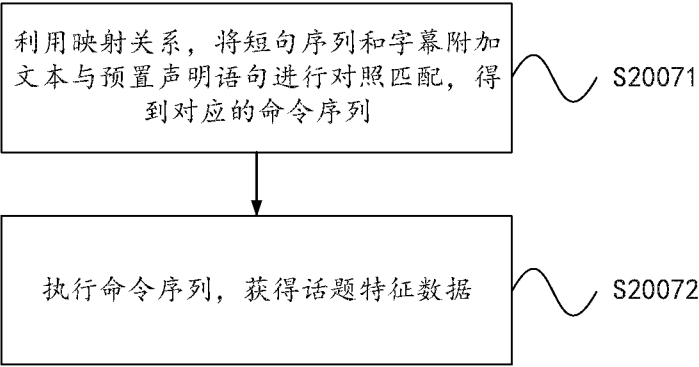


图 4C

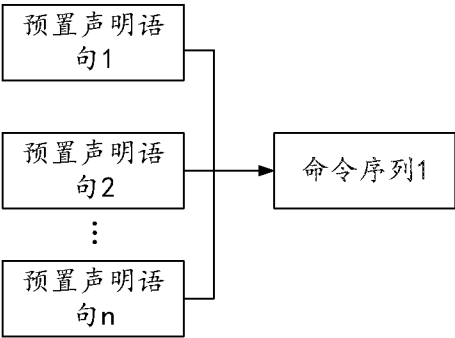


图 4D

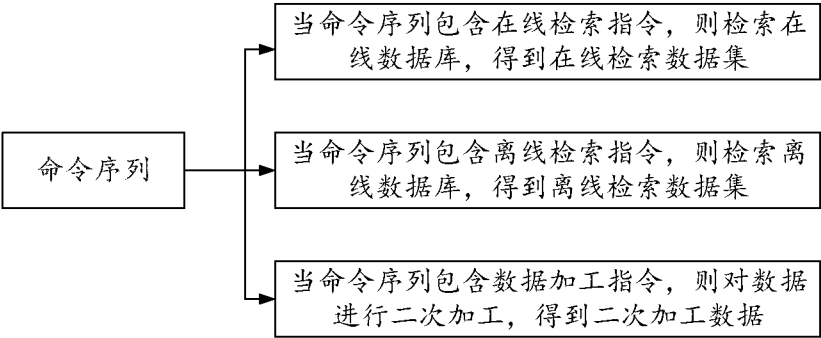


图 4E

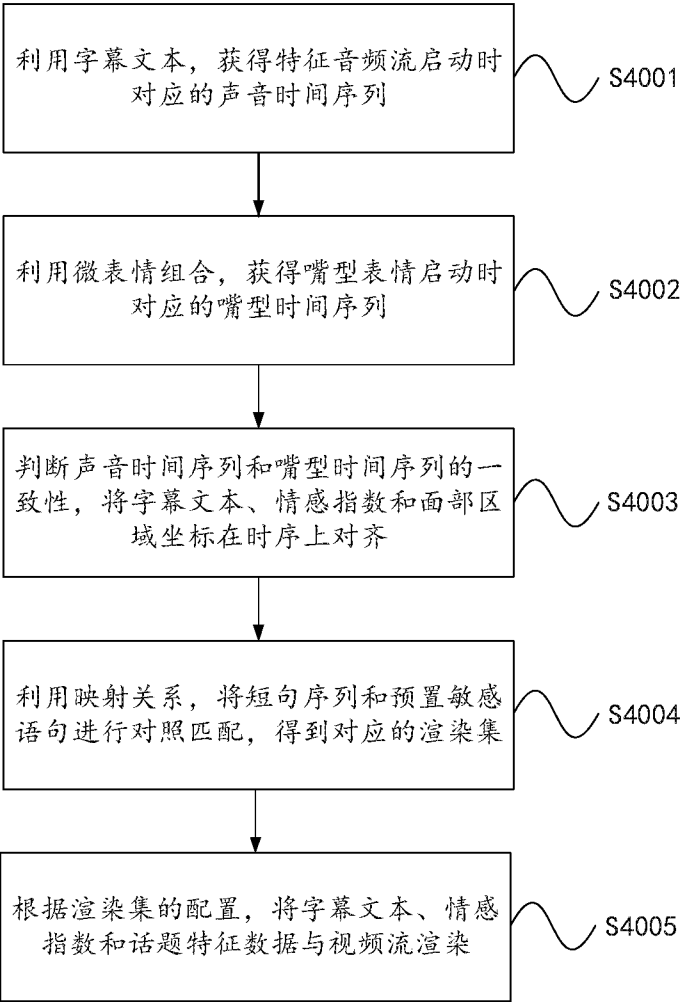


图 5A

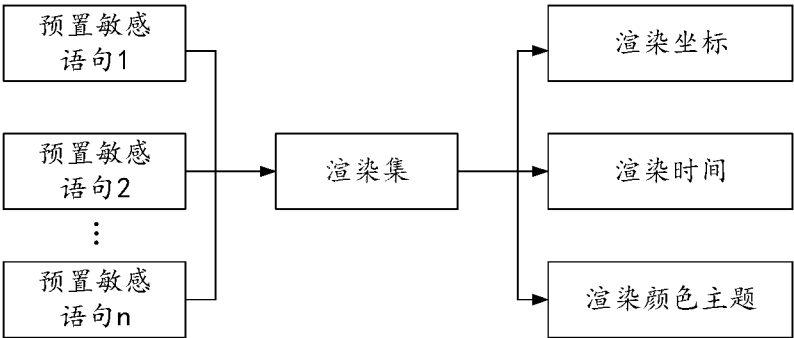


图 5B

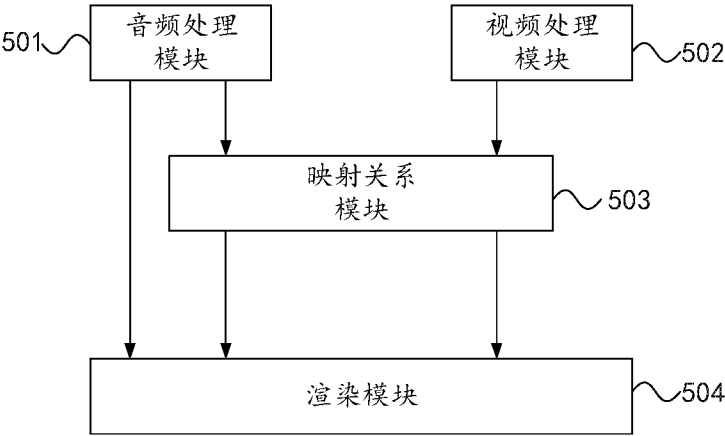


图 6

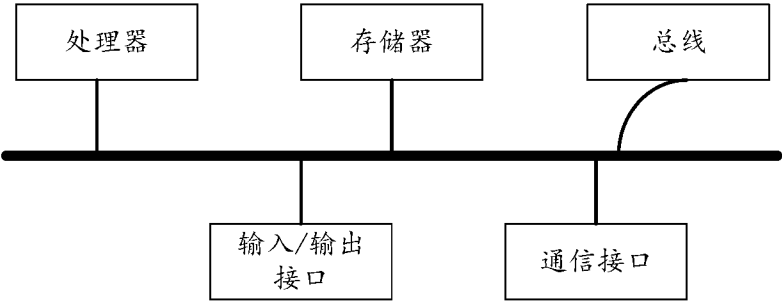


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/101786

A. CLASSIFICATION OF SUBJECT MATTER

H04N 21/233(2011.01)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)

IPC:H04N

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; CNTXT; ENTXT; VEN; USTXT; WPABS; WOTXT; ENTXT; CNKI; CSDN: 视频, 媒体, 直播, 语音, 声音, 音频, 识别, 关联信息, 表情, 情绪, 情感, 显示, video, media, live, sound, audio, recognize, relate, information, face, expression, feel, show

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 109886258 A (XINHUANET (BEIJING) TECHNOLOGY CO., LTD.) 14 June 2019 (2019-06-14) description, paragraphs 2-143, and figures 1-11	1-13
Y	CN 110414465 A (BEIJING SENSING TECHNOLOGY CO., LTD.) 05 November 2019 (2019-11-05) description, paragraphs 2-39	1-13
Y	CN 106161873 A (TVM (BEIJING) TECHNOLOGY CO., LTD.) 23 November 2016 (2016-11-23) description, paragraphs 48-108	1-13
A	CN 103365959 A (SHENZHEN AIDUFEI TECHNOLOGY CO., LTD.) 23 October 2013 (2013-10-23) entire document	1-13
A	CN 110365933 A (WUHAN XINGTU XINKE ELECTRONIC CO., LTD.) 22 October 2019 (2019-10-22) entire document	1-13



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

“D” document cited by the applicant in the international application

“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

16 September 2023

Date of mailing of the international search report

19 September 2023

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/101786

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 114095782 A (GUANGZHOU BOGUAN INFORMATION SCIENCE & TECHNOLOGY CO., LTD.) 25 February 2022 (2022-02-25) entire document	1-13

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/101786

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	109886258	A	14 June 2019	None	
CN	110414465	A	05 November 2019	None	
CN	106161873	A	23 November 2016	None	
CN	103365959	A	23 October 2013	None	
CN	110365933	A	22 October 2019	None	
CN	114095782	A	25 February 2022	None	

国际检索报告

国际申请号

PCT/CN2023/101786

A. 主题的分类

H04N 21/233 (2011.01) 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)

IPC:H04N

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; CNTXT; ENTXT; VEN; USTXT; WPABS; WOTXT; ENTXT; CNKI; CSDN: 视频, 媒体, 直播, 语音, 声音, 音频, 识别, 关联信息, 表情, 情绪, 情感, 显示, video, media, live, sound, audio, recognize, relate, information, face, expression, feel, show

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 109886258 A (新华网 (北京) 科技有限公司) 2019年6月14日 (2019 - 06 - 14) 说明书第2段—第143段, 图1—图11	1-13
Y	CN 110414465 A (北京深醒科技有限公司) 2019年11月5日 (2019 - 11 - 05) 说明书第2段—第39段	1-13
Y	CN 106161873 A (天脉聚源 (北京) 科技有限公司) 2016年11月23日 (2016 - 11 - 23) 说明书第48段—第108段	1-13
A	CN 103365959 A (深圳市爱渡飞科技有限公司) 2013年10月23日 (2013 - 10 - 23) 全文	1-13
A	CN 110365933 A (武汉兴图新科电子股份有限公司) 2019年10月22日 (2019 - 10 - 22) 全文	1-13
A	CN 114095782 A (广州博冠信息科技有限公司) 2022年2月25日 (2022 - 02 - 25) 全文	1-13

☐ 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

“D” 申请人在国际申请中引证的文件

“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

2023年9月16日

国际检索报告邮寄日期

2023年9月19日

ISA/CN的名称和邮寄地址

中国国家知识产权局

中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

魏雨农

电话号码 (+86) 010-62411589

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/101786

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	109886258	A	2019年6月14日	无	
CN	110414465	A	2019年11月5日	无	
CN	106161873	A	2016年11月23日	无	
CN	103365959	A	2013年10月23日	无	
CN	110365933	A	2019年10月22日	无	
CN	114095782	A	2022年2月25日	无	