

# (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  
2014 年 5 月 30 日 (30.05.2014)



(10) 国际公布号  
WO 2014/079324 A1

- (51) 国际专利分类号:  
G06F 3/16 (2006.01) G10L 15/02 (2006.01)
- (21) 国际申请号: PCT/CN2013/086734
- (22) 国际申请日: 2013 年 11 月 8 日 (08.11.2013)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
201210487130.3 2012 年 11 月 26 日 (26.11.2012) CN
- (71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518000 (CN)。
- (72) 发明人: 周彬 (ZHOU, Bin); 中国广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室, Guangdong 518000 (CN)。
- (74) 代理人: 北京德琦知识产权代理有限公司 (DEQI INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区知春路 1 号学院国际大厦 7 层, Beijing 100083 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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)。

## 根据细则 4.17 的声明:

- 关于申请人有权要求在先申请的优先权(细则 4.17(iii))

## 本国际公布:

- 包括国际检索报告(条约第 21 条(3))。

(54) Title: VOICE INTERACTION METHOD AND DEVICE

(54) 发明名称: 一种语音交互方法和装置

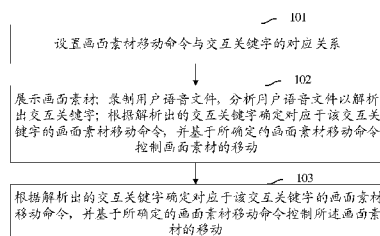


图 1 /Fig. 1

101 Set correlation between picture material movement command and interaction keyword and  
102 Display picture material; record user voice file, and analyze the user voice file so as to parse out the interaction keyword  
103 Determine, according to parsed out interaction keyword, picture material movement command corresponding to interaction keyword, and control movement of picture material based on determined picture material movement command.

(57) Abstract: A voice interaction method and device. The method comprises: setting a correlation between a picture material movement command and an interaction keyword, and the method also comprises: displaying a picture material; recording a user voice file, and analyzing the user voice file so as to parse out the interaction keyword; and according to the parsed out interaction keyword, determining the picture material movement command corresponding to the interaction keyword, and controlling the movement of the picture material based on the determined picture material movement command.

(57) 摘要: 一种语音交互方法和装置, 方法包括: 设置画面素材移动命令与交互关键字的对应关系, 该方法还包括: 展示画面素材; 录制用户语音文件, 分析用户语音文件以解析出交互关键字; 根据解析出的交互关键字确定对应于交互关键字的画面素材移动命令, 并基于所确定的画面素材移动命令控制画面素材的移动。

WO 2014/079324 A1

## 一种语音交互方法和装置

### 技术领域

本发明涉及信息处理技术领域，更具体地，涉及一种语音交互方法和装置。

### 5 发明背景

随着计算机技术和网络技术的飞速发展，互联网（Internet）和即时通信技术在人们的日常生活、学习和工作中发挥的作用也越来越大。而且，随着移动互联网的发展，互联网也在向移动化发展。

当今社会已经进入了高度发达的信息时代，其企业竞争形态也已经  
10 由过去主要以产品功能质量为主的单一竞争形态转变为以企业形象、商品、品牌等作为重要手段和主要倾向的复合竞争形态。这种形态的转变与现代画面展示事业（例如广告）的迅速发展是分不开的。

在目前的画面展示方法中，通常由画面提供商自己直接提供画面素材，而且单向在网络上主动展示画面。

### 15 发明内容

本发明实施方式提出一种语音交互方法，以提高交互成功率。

本发明实施方式还提出了一种语音交互装置，以提高交互成功率。

本发明实施方式还提出了一种移动终端，以提高交互成功率。

本发明实施方式的具体方案如下：

20 一种语音交互方法，设置画面素材移动命令与交互关键字的对应关系，该方法还包括：

展示画面素材；

录制用户语音文件，分析所述用户语音文件以解析出交互关键字；

根据解析出的所述交互关键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制所述画面素材的移动。

一种语音交互装置，包括一个或者多个处理器和存储器；其中，所述存储器中包含可由所述一个或者多个处理器执行的多个单元，所述多个单元包括：对应关系设置单元、画面素材展示单元、交互关键字解析单元和画面素材移动单元，其中：

对应关系设置单元，用于设置画面素材移动命令与交互关键字的对应关系；

画面素材展示单元，用于展示画面素材；

交互关键字解析单元，用于录制用户语音文件，分析所述用户语音文件以解析出交互关键字；

画面素材移动单元，用于根据解析出的所述交互关键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制所述画面素材的移动。

一种移动终端，该移动终端包括一个或者多个处理器和存储器；其中，所述存储器包含可由所述一个或者多个处理器执行的多个单元，所述多个单元包括：显示单元、语音录制单元和计算单元，其中：

显示单元，用于展示画面素材；

语音录制单元，用于录制用户语音文件；

计算单元，用于保存画面素材移动命令与交互关键字的对应关系，分析所述用户语音文件以解析出交互关键字；根据解析出的所述交互关键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制所述画面素材的移动。

从上述技术方案可以看出，在本发明实施方式中，设置画面素材移

- 动命令与交互关键字的对应关系；展示画面素材；录制用户语音文件，分析用户语音文件以解析出交互关键字；根据解析出的交互关键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制该画面素材的移动。由此可见，应用本发明实施方式之后，
- 5 不同于现有技术中画面提供者的单片面画面展示，画面浏览受众可以基于语音方式控制画面素材的移动，因此画面浏览受众可以通过语音方式与画面素材有效交互，提高了交互成功率。

而且，本发明实施方式通过感应用户语音来控制画面素材，同时还提高了画面素材的曝光程度，从而进一步能够提高画面素材投放效果。

## 10 附图简要说明

- 图 1 为本发明实施例中的一种语音交互方法流程图。
- 图 2 为本发明实施例中的一种汽车类型的画面素材示意图。
- 图 3 为本发明实施例中的一种汽车类型的画面素材交互移动示意图。
- 图 4 为本发明实施例中的一种语音交互装置结构图。
- 15 图 5 为本发明实施例中的另一种语音交互装置的结构图。
- 图 6 为本发明实施例中的一种移动终端结构示意图。
- 图 7 为本发明实施例中的另一种移动终端的结构示意图。

## 具体实施方式

- 为使本发明的目的、技术方案和优点更加清楚，下面结合附图对本
- 20 发明作进一步的详细阐述。

在目前的画面展示方法中，通常由画面提供商自己直接提供画面素材，而且单向在网络上主动展示画面。然而，这种展示方式并没有考虑到画面受众的个人参与程度，因此仅是一种片面的画面展示，与

画面浏览受众缺乏有效交互，因此交互成功率很低。

为此，本发明实施例提供了一种语音交互方法。在本发明实施方式中，设置画面素材移动命令与交互关键字的对应关系；展示画面素材；录制用户语音文件，分析用户语音文件以解析出交互关键字；根据解析出的交互关键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制画面素材的移动。由此可见，应用本发明实施方式之后，不同于现有技术中画面提供者的单片面画面展示，画面浏览受众可以基于语音方式控制画面素材的移动，因此画面浏览受众可以通过语音方式与画面素材有效交互，提高了交互成功率。

图 1 为根据本发明实施方式的语音交互方法流程图。

如图 1 所示，该方法包括：

步骤 101：设置画面素材移动命令与交互关键字的对应关系。

在这里，画面素材移动命令用于控制画面素材的移动。可以设置画面素材加速命令、画面素材减速命令、画面素材启动命令、画面素材停止命令、画面素材移动速度保持命令、或画面素材移动轨迹等各种画面素材移动命令与交互关键字的对应关系。

当用户的语音中包含交互关键字时，可以基于与该交互关键字对应的画面素材移动命令来控制画面素材的移动。比如，可以设置交互关键字“启动”对应于画面素材启动命令；交互关键字“停止”对应于画面素材停止命令；设置交互关键字“加速”对应于画面素材加速命令；设置交互关键字“减速”对应于画面素材减速命令；交互关键字“曲线”对应于设置画面素材移动轨迹为曲线的命令；交互关键字“直线”对应于设置画面素材移动轨迹为直线的命令，等等。

以上虽然详细罗列了画面素材移动命令与交互关键字的具体实例，本领域技术人员可以意识到，这种罗列仅仅是示范性的，并不用于限定

本发明实施方式的保护范围。

在一个实施方式中，可以将画面素材移动速度与交互关键字的对应关系保存在移动终端上。移动终端可以包括但是不局限于：功能手机、智能手机、掌上电脑、个人电脑(PC)、平板电脑或个人数字助理(PDA)，  
5 等等。

以上虽然详细罗列了移动终端的具体实例，本领域人员可以意识到，这些罗列仅是阐述目的，并不用于限定本发明实施方式的保护范围。

移动终端可以适用于任意的智能终端操作系统，具体可以采用的操作系统包括但是不局限于：安卓(Android)、Palm OS、Symbian(塞班)、  
10 Windows mobile、Linux、Android(安卓)、iPhone(苹果) OS、Black Berry (黑莓) OS 6.0、Windows Phone 系列，等等。

优选地，移动终端具体可以采用安卓操作系统，而且移动终端可以采用到 Android 的各个版本之中，包括但是不局限于：阿童木(Android Beta)、发条机器人(Android 1.0)、纸杯蛋糕(Android 1.5)、甜甜圈  
15 (Android 1.6)、松饼(Android 2.0/2.1)、冻酸奶(Android 2.2)、姜饼(Android 2.3)、蜂巢(Android 3.0)、冰激凌三明治(Android 4.0)、果冻豆(Jelly Bean, Android 4.1)等版本。以上详细罗列了 Android 平台的具体版本，本领域技术人员可以意识到，本发明实施方式并不局限于上述罗列版本，而还可以适用于其他基于 Android 软件架构的任意版本  
20 之中。

需要注意的是，在设置完画面素材移动命令和交互关键字的对应关系之后，当用户下次利用语音控制画面素材的移动时，可以直接执行步骤 102，步骤 101 可以不再重复执行。

步骤 102：展示画面素材，录制用户语音文件，分析所述用户语音  
25 文件以解析出交互关键字。

在这里，图像受众、图像提供商或任意的第三方都可以通过各种信息传输方式，将画面素材上传到位于网络侧的服务器中，然后由移动终端从服务器中获取到画面素材，并展示画面素材。

画面素材的具体内容与期望向用户展示的素材本身相关。比如，假设期望投放关于品牌汽车的信息推送，则可以上传品牌汽车的实物模型图像；假如期望投放关于某款电子产品的信息推送，则可以上传该款电子产品的实物模型图像；等等。

在移动终端上还可以展示与画面素材相对应的背景图像。背景图像和画面素材的种类可以有多种。比如，背景图像和画面素材可以分别为：位图；联合照片专家组（JPEG）；签图像文件格式（TIFF）；图像互换格式（GIF）；流式网络图形格式（PNG）；或三维图像，等等。GIF 图像文件的数据是经过压缩的，而且是采用了可变长度等压缩算法。GIF 格式的另一个特点是其在一个 GIF 文件中可以存多幅彩色图像，如果把存于一个文件中的多幅图像数据逐幅读出并显示到屏幕上，就可构成一种最简单的画面。

本发明实施方式中，可以通过将该画面类型的画面素材叠加到背景图像上，实现共同显示。

示范性地，图 2 为根据本发明实施方式的汽车类型的画面素材示意图。由图 2 可见，画面素材为汽车模型，背景图像为房屋。汽车模型与房屋融合显示在一起，而且汽车模型停留在房屋之中。

而且，信息推送受众、图像提供商或任意的第三方还可以通过多种方式上传画面素材。比如，可以由图像受众直接在移动终端上，通过无线上网的方式将画面素材上传到服务器；也可以由图像提供商通过个人计算机（PC）等方式将画面素材上传到服务器，等等。

优选的，移动终端可以通过通用网关接口（CGI）的方式从服务器

获取到背景图像和画面素材，并在移动终端的浏览器窗口上显示背景图像和画面素材。

移动终端浏览器是运行在移动终端上的浏览器，可以通过通用分组无线服务技术（GPRS）等多种方式上网浏览互联网内容。目前，一些移动终端浏览器需要 JAVA 或移动终端的系统（如苹果的 IOS 系统以及 Android 平台等）支持。

服务器可以向移动终端提供背景图像，也可以在移动终端本地预先存储背景图像。优选在服务器中保存背景图像，这是因为相比较移动终端，服务器可以具有更大的储存空间，因此可以在服务器中保存海量的背景图像。示范性地，背景图像可以包括：蓝天、白云、大桥、公路等等。

在一个实施方式中，图像受众、图像提供商或任意第三方除了上传画面素材之外，还进一步上传用于描述该画面素材类型的图像属性信息。图像属性信息可以是文本信息。还可以直接通过画面素材的命名来描述其类型。服务器可以根据图像属性信息确定出画面素材的类型，并检索与该画面素材相匹配的背景图像。

比如，假设画面素材是针对汽车的信息推送，则服务器可以检索出适于汽车的背景图像（比如为跑道）；假设画面素材是针对电子产品的信息推送，则服务器可以检索出适于电子产品的背景图像（比如为办公桌）。

在这里，当由服务器提供背景图像时，服务器可以首先将上传来的画面素材和自身保存的背景图像发送到移动终端，而且优选顺带发送定单信息和/或广告位信息等，并提供给移动终端进行相应展示。

在移动终端上将背景图像和画面素材共同显示，从而实现融合显示。优选地，画面素材显示在背景图像的上部或前面。

当向用户展示画面素材之后，可以基于用户的语音实现针对画面素材的互动过程。

用户在浏览画面素材时，或者浏览画面素材之后，可以发出语音。此时录制用户语音文件，分析用户语音文件以解析出交互关键字。

- 5        在这里，可以基于各种语音识别技术分析用户语音文件以解析出交互关键字。语音识别技术主要包括特征提取技术、模式匹配准则及模型训练技术等方面。

本发明实施方式可以采用的语音识别的种类有很多，比如连续语音识别、关键词检出、说话人辨认、说话人确认、语音合成、音频检索等方式。更具体地，连续语音识别可以采用隐马尔科夫模型，而且本发明实施方式还可以采用动态时间规正、神经网络、支持向量机、矢量量化等各种语音识别算法。

在具体实施中，可以将各种语音识别技术通过内置插件或接口等方式嵌入到各种终端的浏览器窗口中，从而使得浏览器窗口自身可以具备  
15        相应语音识别功能。

比如，可以将用户输入的语音文件转换成文本文件，再将文本文件与数据库中的文本格式的关键字进行比较，如果匹配成功则可确定出交互关键字。也可以将用户输入的语音文件与语音格式的交互关键字进行波形比较，若与语音格式的交互关键字波形一致，则可确定出交互关键字。  
20        字。

在一个实施方式中，可以首先获取语音训练文件和文本训练文件，并利用所述语音训练文件和文本训练文件，对语音识别器采用语音自适应方式估计语音参数，并利用该已估计语音参数的语音识别器识别所述用户语音文件，以将该用户语音文件转换为文本文件；再从所述文本文件  
25        件中检索交互关键字。

在另一个实施方式中，可以确定该用户语音文件的语音波形图；判断该用户语音文件的语音波形图中是否包含与交互关键字的语音波形保持一致的波形区域，如果是，则基于该所包含的交互关键字的语音波形确定该交互关键字。

5       在本发明实施方式中，还可以基于语音来实现变更画面素材。

具体地，可以设置画面素材变更命令与画面素材变更关键字的对应关系；然后通过录制用户语音文件，分析所述用户语音文件以解析出画面素材变更关键字；再根据解析出的所述画面素材变更关键字确定对应于该画面素材变更关键字的更新后画面素材，并展示更新后画面素材。

10       步骤 103：根据解析出的所述交互关键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制所述画面素材的移动。

在这里，可以根据在步骤 101 设置的画面素材移动命令与交互关键字的对应关系，根据解析出的所述交互关键字确定对应于该交互关键字  
15       的画面素材移动命令。

比如，假如在步骤 101 中设置有关键字“启动”对应于画面素材启动命令、交互关键字“停止”对应于画面素材停止命令、设置交互关键字“加速”对应于画面素材加速命令、设置交互关键字“减速”对应于画面素材减速命令、交互关键字“曲线”对应于设置画面素材移动轨迹  
20       为曲线的命令、交互关键字“直线”对应于设置画面素材移动轨迹为直线的命令。

则当步骤 102 中解析出交互关键字为“启动”时，则可确定画面素材移动命令具体为：画面素材启动命令；当步骤 102 中解析出交互关键字为“停止”时，则可确定画面素材移动命令具体为：画面素材停止命令；当步骤 102 中解析出交互关键字为“加速”时，则可确定画面素材  
25

移动命令具体为：画面素材加速命令；当步骤 102 中解析出交互关键字为“减速”时，则可确定画面素材移动命令具体为：画面素材减速命令；当步骤 102 中解析出交互关键字为“曲线”时，则可确定画面素材移动命令具体为：设置画面素材移动轨迹为曲线的命令；当步骤 102 中解析出交互关键字为“直线”时，则可确定画面素材移动命令具体为：设置画面素材移动轨迹为直线的命令，等等。

当解析出的所述交互关键字确定对应于该交互关键字的画面素材移动命令，可以基于所确定的画面素材移动命令控制所述画面素材的移动。

在本发明实施方式中，还可以在步骤 101 中预先设置交互关键字与画面素材移动速度之间的对应关系。比如：设置交互关键字“高速运动”对应于画面素材保持预先设置的高速移动速度，即针对画面素材的高速移动速度保持命令；还可以设置交互关键字“中速运动”对应于画面素材保持为预先设置的中速移动速度，即针对画面素材的中速移动速度保持命令；还可以设置交互关键字“低速运动”对应于画面素材保持为预先设置的低速移动速度，即针对画面素材的低速移动速度保持命令。

当步骤 102 中解析出交互关键字为“高速运动”时，则可确定画面素材移动命令具体为：画面素材保持预先设置的高速移动速度，则此时控制画面素材移动速度加速到该预先设置的高速移动速度，并保持该高速移动速度。

当步骤 102 中解析出交互关键字为“中速运动”时，则可确定画面素材移动命令具体为：画面素材保持预先设置的中速移动速度，则此时控制画面素材移动速度加速到该预先设置的中速移动速度，并保持该中速移动速度。

当步骤 102 中解析出交互关键字为“低速运动”时，则可确定画面素材移动命令具体为：画面素材保持预先设置的低速移动速度，则此时

控制画面素材移动速度加速到该预先设置的低速移动速度，并保持该低速移动速度。

图 3 为根据本发明实施方式的汽车类型的画面素材交互移动示意图。

由图 3 可见，画面素材为汽车模型，背景图像为房屋。汽车模型与房屋融合显示在一起，而且汽车模型停留在房屋之上。当感应到用户的语音之后，可以根据包含于该语音中的交互关键字对应的画面素材移动命令，汽车模型从该房屋中移开。

而且，在用户靠近摄像头时，可以由摄像头的热传感器检测到用户的热量，并开始播放预先设定的引擎启动声音，以预示汽车模型已经启动。

基于上述详细分析，本发明实施方式还提出了一种语音交互装置。

图 4 为根据本发明实施方式的语音交互装置结构图。

如图 4 所示，该装置包括对应关系设置单元 401、画面素材展示单元 402、交互关键字解析单元 403 和画面素材移动单元 404，其中：

对应关系设置单元 401，用于设置画面素材移动命令与交互关键字的对应关系；其中，画面素材移动命令用于控制画面素材的移动；

画面素材单元 402，用于展示画面素材；

交互关键字解析单元 403，用于录制用户语音文件，分析所述用户语音文件以解析出交互关键字；

画面素材移动单元 404，用于根据解析出的所述交互关键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制所述画面素材的移动。

在一个实施方式中，交互关键字解析单元 403，用于获取语音训练文件和文本训练文件，并利用所述语音训练文件和文本训练文件，对语音识别器采用语音自适应方式估计语音参数，并利用该已估计语音参数

的语音识别器识别所述用户语音文件，以将该用户语音文件转换为文本文件；从所述文本文件中检索交互关键字。

在另一个实施方式中，交互关键字解析单元 403，用于确定该用户语音文件的语音波形图；判断该用户语音文件的语音波形图中是否包含与交互关键字的语音波形保持一致的波形区域，如果是，则基于该所包含的交互关键字的语音波形确定该交互关键字。

优选地，对应关系设置单元 401，用于设置画面素材加速命令、画面素材减速命令、画面素材启动命令、画面素材停止命令、画面素材移动速度保持命令、或画面素材移动轨迹与交互关键字的对应关系。

图 5 为本发明实施例中另一种语音交互装置的结构示意图。如图 5 所示。除了对应关系设置单元 401、画面素材展示单元 402、交互关键字解析单元 403 和画面素材移动单元 404，该装置可以进一步包括画面素材变更单元 405；其中，

对应关系设置单元 401，进一步用于设置画面素材变更命令与画面素材变更关键字的对应关系；

交互关键字解析单元 403，进一步用于录制用户语音文件，分析所述用户语音文件以解析出画面素材变更关键字；

画面素材变更单元 405，用于根据解析出的所述画面素材变更关键字确定对应于该画面素材变更关键字的更新后画面素材，并向画面素材展示单元 402 发送该更新后画面素材；

画面素材展示单元 402，进一步用于展示该更新后画面素材。

可以基于上述方法和装置，在多种应用环境中实施本发明实施方式。比如，可以将本发明实施方式应用到移动终端的广告应用中。可以预先设置互动关键字“奥迪启动”。在用户点击广告后，提示用户利用麦克风说提示语：“奥迪启动”；当用户说出奥迪启动后，该语音关键字被解

析成文本，与广告设定的互动关键字比较。若互动关键字一致，则将触发广告的行为，放出汽车启动的引擎声音，预示着广告内的汽车已经启动，而且广告横幅内的汽车轮胎转动，迅速移动开出广告可见范围，因此更能提高广告的互动性和新颖性。

5       基于上述分析，本发明实施方式还提出了一种移动终端。

图 6 为根据本发明实施方式的移动终端结构示意图。

如图 6 所示，该移动终端包括显示单元 601、语音录制单元 602 和计算单元 603。其中：

显示单元 601，用于展示画面素材；

10       语音录制单元 602，用于录制用户语音文件；

计算单元 603，用于保存画面素材移动命令与交互关键字的对应关系，分析所述用户语音文件以解析出交互关键字；根据解析出的所述交互关键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制所述画面素材的移动。

15       图 7 为本发明实施例中另一种移动终端的结构示意图。如图 7 所示，除了显示单元 601、语音录制单元 602 和计算单元 603，该移动终端可以进一步包括摄像单元 604，用于感应用户的热量，并在感应到用户热量后向显示单元 601 发出热量提示消息；

显示单元 601，进一步用于收到热量提示消息之后，播放画面素材  
20       移动启动音频。

本发明实施例中的单元可以集成于一体，也可以分离部署；可以合并为一个单元，也可以进一步拆分成多个子单元。这些单元可以由软件实现（如存储在非易失性存储介质中的由一个或者多个处理器执行的计算机可读指令），或者由硬件实现，或者由软件和硬件结合实现。

25       而且，在这里，图像受众、图像提供商或任意的第三方都可以通过

各种信息传输方式，将画面素材上传到位于网络侧的服务器中，然后由移动终端从服务器中获取到画面素材，并展示画面素材。

实际上，可以通过多种形式来具体实施本发明实施方式所提出的语音互动方法和装置。比如，可以遵循一定规范的应用程序接口，将语音互动方法编写为安装到移动终端中的插件程序，也可以将其封装为应用程序以供用户自行下载使用。当编写为插件程序时，可以将其实施为 ocx、dll、cab 等多种插件形式。也可以通过 Flash 插件、RealPlayer 插件、MMS 插件、MIDI 五线谱插件、ActiveX 插件等具体技术来实施本发明实施方式所提出的语音互动方法。

可以通过指令或指令集存储的储存方式将本发明实施方式所提出的语音互动方法存储在各种存储介质上。这些存储介质包括但不限于：软盘、光盘、DVD、硬盘、闪存、U 盘、CF 卡、SD 卡、MMC 卡、SM 卡、记忆棒（Memory Stick）、xD 卡等。

另外，还可以将本发明实施方式所提出的语音互动方法应用到基于闪存（Nand flash）的存储介质中，比如 U 盘、CF 卡、SD 卡、SDHC 卡、MMC 卡、SM 卡、记忆棒、xD 卡等。

综上所述，在本发明实施方式中，设置画面素材移动命令与交互关键字的对应关系，该方法还包括：展示画面素材；制用户语音文件，分析所述用户语音文件以解析出交互关键字；根据解析出的所述交互关键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制所述画面素材的移动。由此可见，应用本发明实施方式之后，不同于现有技术中画面提供者的单片面画面展示，而是可以基于语音文件的交互关键字控制画面素材的移动，因此画面浏览受众可以通过语音方式与画面素材有效交互，提高了交互成功率。

而且，本发明实施方式通过感应用户语音来控制画面素材，同时还

提高了画面素材的曝光程度，从而进一步能够提高画面素材投放效果。

以上所述，仅为本发明的较佳实施例而已，并非用于限定本发明的保护范围。凡在本发明的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

- 5       以上所述仅为本发明的较佳实施例而已，并不用以限制本发明，凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等，均应包含在本发明的保护范围之内。

## 权利要求书

1、一种语音交互方法，其特征在于，设置画面素材移动命令与交互关键字的对应关系，该方法还包括：

展示画面素材；

5        录制用户语音文件，分析所述用户语音文件以解析出交互关键字；

根据解析出的所述交互关键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制所述画面素材的移动。

2、根据权利要求1所述的语音交互方法，其特征在于，所述分析用户语音文件以解析出交互关键字包括：

10       获取语音训练文件和文本训练文件，并利用所述语音训练文件和文本训练文件，对语音识别器采用语音自适应方式估计语音参数，并利用该已估计语音参数的语音识别器识别所述用户语音文件，以将该用户语音文件转换为文本文件；

从所述文本文件中检索交互关键字。

15       3、根据权利要求1所述的语音交互方法，其特征在于，所述分析用户语音文件以解析出交互关键字包括：

确定该用户语音文件的语音波形图；

判断该用户语音文件的语音波形图中是否包含与交互关键字的语音波形保持一致的波形区域，如果是，则基于该所包含的交互关键字的语音波形确定该交互关键字。

20       4、根据权利要求1所述的语音交互方法，其特征在于，所述设置画面素材移动命令与交互关键字的对应关系为：设置画面素材加速命令、画面素材减速命令、画面素材启动命令、画面素材停止命令、画面素材移动速度保持命令、或画面素材移动轨迹与交互关键字的对应关系。

5、根据权利要求1所述的语音交互方法，其特征在于，进一步设置画面素材变更命令与画面素材变更关键字的对应关系；该方法还包括：

录制用户语音文件，分析所述用户语音文件以解析出画面素材变更关键字；

5 根据解析出的所述画面素材变更关键字确定对应于该画面素材变更关键字的更新后画面素材，并展示所述更新后画面素材。

6、一种语音交互装置，其特征在于，包括一个或者多个处理器和存储器；其中，所述存储器中包含可由所述一个或者多个处理器执行的多个单元，所述多个单元包括：对应关系设置单元、画面素材展示单元、

10 交互关键字解析单元和画面素材移动单元，其中：

对应关系设置单元，用于设置画面素材移动命令与交互关键字的对应关系；

画面素材展示单元，用于展示画面素材；

交互关键字解析单元，用于录制用户语音文件，分析所述用户语音  
15 文件以解析出交互关键字；

画面素材移动单元，用于根据解析出的所述交互关键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制所述画面素材的移动。

7、根据权利要求6所述的语音交互装置，其特征在于，

20 交互关键字解析单元，用于获取语音训练文件和文本训练文件，并利用所述语音训练文件和文本训练文件，对语音识别器采用语音自适应方式估计语音参数，并利用该已估计语音参数的语音识别器识别所述用户语音文件，以将该用户语音文件转换为文本文件；从所述文本文件中检索交互关键字。

25 8、根据权利要求6所述的语音交互装置，其特征在于，

交互关键字解析单元，用于确定该用户语音文件的语音波形图；判断该用户语音文件的语音波形图中是否包含与交互关键字的语音波形保持一致的波形区域，如果是，则基于该所包含的交互关键字的语音波形确定该交互关键字。

5        9、根据权利要求6所述的语音交互装置，其特征在于，

对应关系设置单元，用于设置画面素材加速命令、画面素材减速命令、画面素材启动命令、画面素材停止命令、画面素材移动速度保持命令、或画面素材移动轨迹与交互关键字的对应关系。

10       10、根据权利要求6所述的语音交互装置，其特征在于，进一步包括画面素材变更单元；

对应关系设置单元，进一步用于设置画面素材变更命令与画面素材变更关键字的对应关系；

交互关键字解析单元，进一步用于录制用户语音文件，分析所述用户语音文件以解析出画面素材变更关键字；

15       画面素材变更单元，用于根据解析出的所述画面素材变更关键字确定对应于该画面素材变更关键字的更新后画面素材，并向画面素材展示单元发送该更新后画面素材；

画面素材展示单元，进一步用于展示该更新后画面素材。

20       11、一种移动终端，其特征在于，包括一个或者多个处理器和存储器；其中，所述存储器包含可由所述一个或者多个处理器执行的多个单元，所述多个单元包括：显示单元、语音录制单元和计算单元，其中：

显示单元，用于展示画面素材；

语音录制单元，用于录制用户语音文件；

计算单元，用于保存画面素材移动命令与交互关键字的对应关系，

25       分析所述用户语音文件以解析出交互关键字；根据解析出的所述交互关

键字确定对应于该交互关键字的画面素材移动命令，并基于所确定的画面素材移动命令控制所述画面素材的移动。

12、根据权利要求 11 所述的移动终端，其特征在于，所述计算单元进一步用于：

- 5        获取语音训练文件和文本训练文件，并利用所述语音训练文件和文本训练文件，对语音识别器采用语音自适应方式估计语音参数，并利用该已估计语音参数的语音识别器识别所述用户语音文件，以将该用户语音文件转换为文本文件；

从所述文本文件中检索交互关键字。

- 10       13、根据权利要求 11 所述的移动终端，其特征在于，所述计算单元进一步用于：

确定该用户语音文件的语音波形图；

- 判断该用户语音文件的语音波形图中是否包含与交互关键字的语音波形保持一致的波形区域，如果是，则基于该所包含的交互关键字的语音波形确定该交互关键字。
- 15

14、根据权利要求 11 所述的移动终端，其特征在于，所述画面素材移动命令与交互关键字的对应关系包括：画面素材加速命令、画面素材减速命令、画面素材启动命令、画面素材停止命令、画面素材移动速度保持命令、或画面素材移动轨迹与交互关键字的对应关系。

- 20       15、根据权利要求 11 所述的移动终端，其特征在于，所述计算单元进一步用于保存画面素材变更命令与画面素材变更关键字的对应关系；  
所述计算单元进一步用于：

分析所述用户语音文件以解析出画面素材变更关键字；

- 根据解析出的所述画面素材变更关键字确定对应于该画面素材变更关键字的更新后画面素材，并控制所述展示单元展示所述更新后画面素
- 25

材。

16、根据权利要求 11 所述的移动终端，进一步包括摄像单元，  
摄像单元，进一步用于感应用户的热量，并在感应到用户热量后向  
显示单元发出热量提示消息；

- 5       显示单元，进一步用于收到热量提示消息之后，播放画面素材移动  
启动音频。

1/4

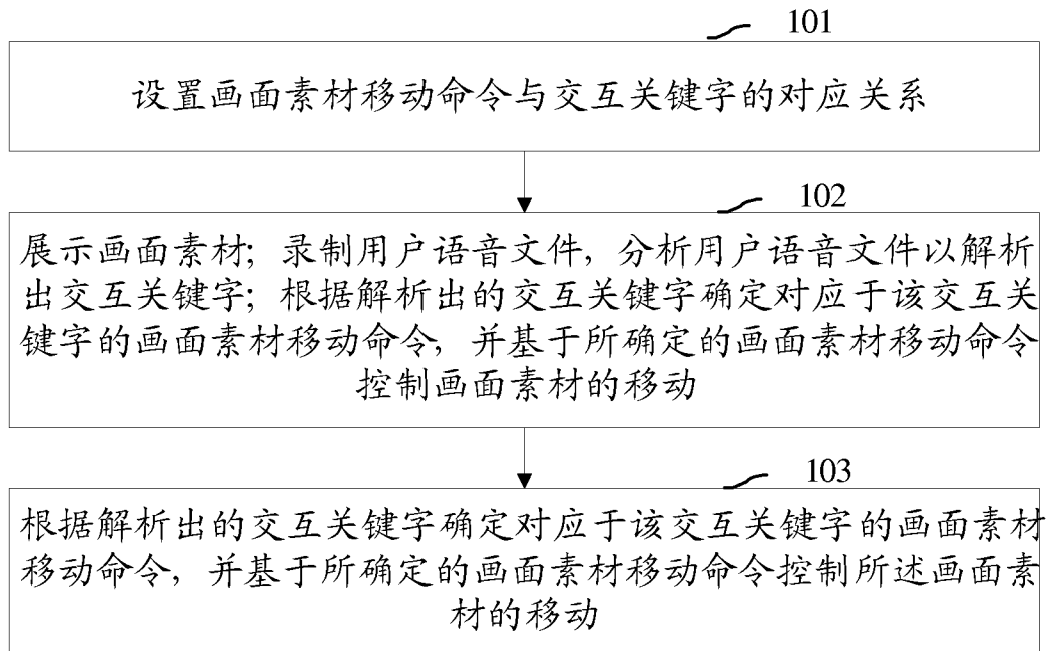


图 1



图 2

2/4

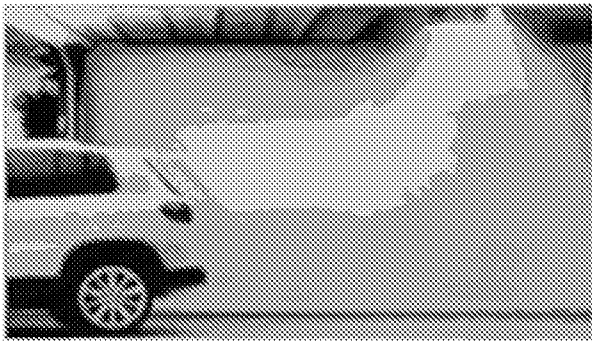


图 3

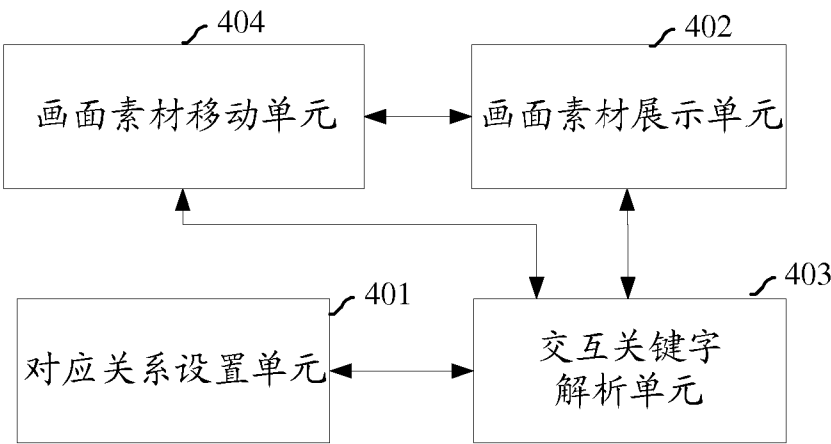


图 4

3/4

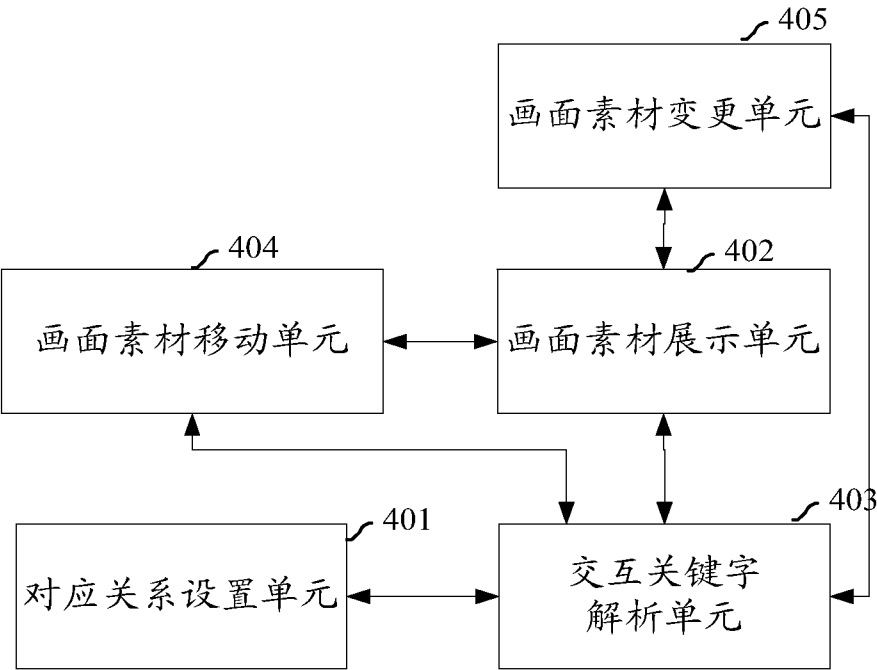


图 5

4/4

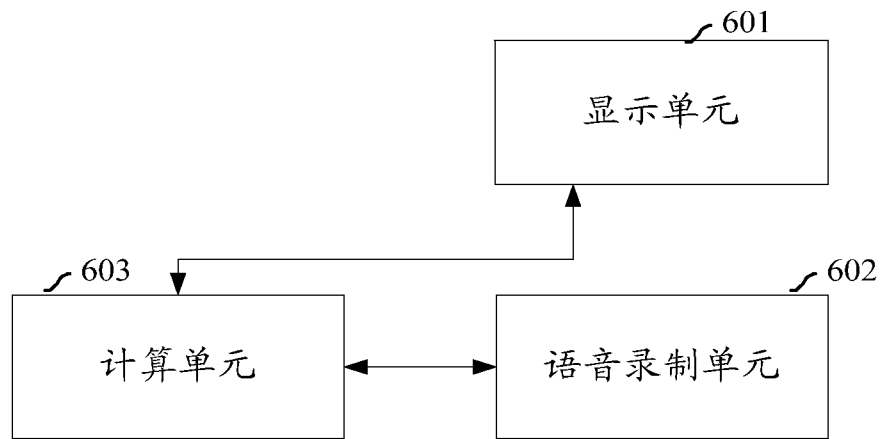


图 6

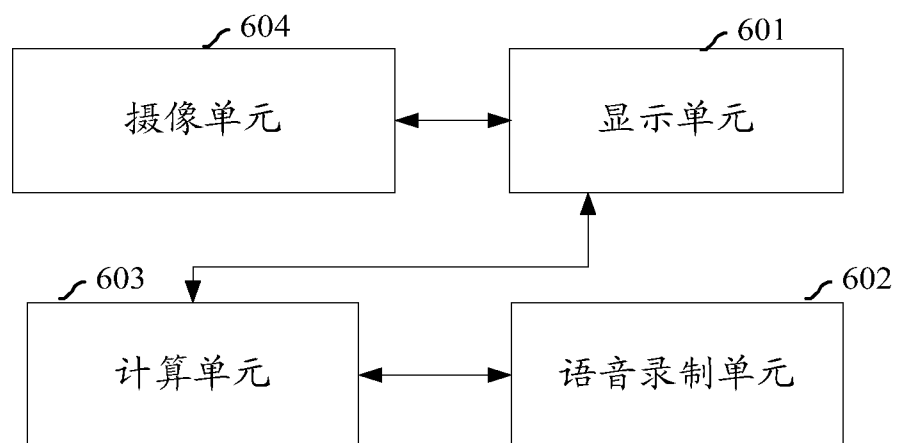


图 7

# INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2013/086734**

## A. CLASSIFICATION OF SUBJECT MATTER

See the extra sheet

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: G06F 3, G10L 15, G10L 17

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, CNKI: picture, material, advertisement, voice, sound, keyword, move, control, interaction, motion, command, instruction, speed, acceleration, deceleration, identification, training, text, camera, heat, temperature, display, play, infrared, automatic  
SIPOABS, DWPI: picture?, imag+, cartoon?, material, animat+, photo+, display+, LCD, screen?, voice, audio, vocal, sound, speech, verbal, interact+, interfac+, key w word?, mov+, speed+, command?, instruct+, order+, camera?, advertis+, ad, heat+, infrared

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                           | Relevant to claim No.             |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
| X         | CN 102253710 A (DELTA ELECTRONICS, INC.), 23 November 2011 (23.11.2011), description, page 4, paragraph [0068] to page 5, paragraph [0077], and figures 1-3                  | 1, 4, 6, 9, 11, 14                |
| Y         |                                                                                                                                                                              | 2, 3, 5, 7, 8, 10, 12, 13, 15, 16 |
| Y         | CN 1371090 A (SUZHOU PEACOCK ELECTRONICS GROUP CO., LTD.), 25 September 2002 (25.09.2002), description, page 5, lines 16-28, and figures 1-2                                 | 2, 7, 12                          |
| Y         | CN 102374864 A (AMBIT MICROSYSTEMS (SHANGHAI) CO., LTD. et al.), 14 March 2012 (14.03.2012), description, page 4, paragraph [0046] to page 5, paragraph [0062], and figure 5 | 3, 8, 13                          |

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

|                                                                                                                                                                         |                                                                                                                                                                                                                                                  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| * Special categories of cited documents:                                                                                                                                | “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                                              |
| “A” document defining the general state of the art which is not considered to be of particular relevance                                                                | “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                                                                     |
| “E” earlier application or patent but published on or after the international filing date                                                                               | “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 |
| “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) | “&” document member of the same patent family                                                                                                                                                                                                    |
| “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                                                                  |                                                                                                                                                                                                                                                  |

|                                                                                                                                                                                                                    |                                                                                           |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| Date of the actual completion of the international search<br>15 January 2014 (15.01.2014)                                                                                                                          | Date of mailing of the international search report<br><b>23 January 2014 (23.01.2014)</b> |
| Name and mailing address of the ISA/CN:<br>State Intellectual Property Office of the P. R. China<br>No. 6, Xitucheng Road, Jimenqiao<br>Haidian District, Beijing 100088, China<br>Facsimile No.: (86-10) 62019451 | Authorized officer<br><b>HE, Xiaolan</b><br>Telephone No.: (86-10) <b>62085135</b>        |

# INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2013/086734**

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages                                                               | Relevant to claim No. |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Y        | US 2011/003777 A1 (APPLE INC.), 17 February 2011 (17.02.2011), description, page 4, paragraph [0041], and page 6, paragraph [0055], and figure 1 | 5, 10, 15             |
| Y        | CN 201117218 Y (ZOU, Peng), 17 September 2008 (17.09.2008), description, page 2, line 18 to page 3, line 11, and figures 1-2                     | 16                    |
| A        | CN 101742110 A (TIANJIN SAMSUNG ELECTRONICS CO., LTD.), 16 June 2010 (16.06.2010), the whole document                                            | 1-16                  |
| A        | JP 2002-318132 A (HITACHI LTD.), 31 October 2002 (31.10.2002), the whole document                                                                | 1-16                  |

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.

**PCT/CN2013/086734**

| Patent Documents referred<br>in the Report | Publication Date | Patent Family      | Publication Date |
|--------------------------------------------|------------------|--------------------|------------------|
| CN 102253710 A                             | 23.11.2011       | None               |                  |
| CN 1371090 A                               | 25.09.2002       | None               |                  |
| CN 102374864 A                             | 14.03.2012       | US 2012/0041766 A1 | 16.02.2012       |
|                                            |                  | US 8412455 B2      | 02.04.2013       |
| US 2011/003777 A1                          | 17.02.2011       | None               |                  |
| CN 201117218 Y                             | 17.09.2008       | None               |                  |
| CN 101742110 A                             | 16.06.2010       | None               |                  |
| JP 2002-318132 A                           | 31.10.2002       | US 2002/0156570 A1 | 24.10.2002       |
|                                            |                  | US 7076362 B2      | 11.07.2006       |
|                                            |                  | DE 60226194D D1    | 05.06.2008       |
|                                            |                  | DE 60226194T T2    | 14.05.2009       |
|                                            |                  | EP 1253569 A2      | 30.10.2002       |
|                                            |                  | EP 1253569B1 B1    | 23.04.2008       |

# INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2013/086734**

## CONTINUATION: CLASSIFICATION OF SUBJECT MATTER

G06F 3/16 (2006.01) i

G10L 15/02 (2013.01) i

**A. 主题的分类**

参见附加页

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: G06F3, G10L15, G10L17

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、CNKI: 画面, 图画, 图片, 素材, 广告, 语音, 声音, 话音, 关键字, 移动, 控制, 交互, 移动, 活动, 命令, 指令, 速度, 加速, 减速, 识别, 训练, 文本, 摄像, 热量, 温度, 显示, 播放, 广告, 红外, 自动

SIPOABS、DWPI: picture?, imag+, cartoon?, material, animat+, photo+, display+, LCD, screen?, voice, audio, vocal, sound, speech, verbal, interact+, interfac+, key w word?, mov+, speed+, command?, instruct+, order+, camera?, advertis+, ad, heat+, infrared

**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                  |
|------|--------------------------------------------------------------------------------------------------|--------------------------|
| X    | CN 102253710 A (台达电子工业股份有限公司) 23.11 月 2011<br>(23.11.2011) 说明书第 4 页第[0068]段—第 5 页第[0077]段以及图 1—3 | 1,4,6,9,11,14            |
| Y    |                                                                                                  | 2,3,5,7,8,10,12,13,15,16 |
| Y    | CN 1371090 A (苏州孔雀电器集团有限责任公司) 25.09 月 2002<br>(25.09.2002) 说明书第 5 页第 16—28 行以及图 1—2              | 2,7,12                   |
| Y    | CN 102374864 A (国基电子(上海)有限公司等) 14.03 月 2012<br>(14.03.2012) 说明书第 4 页第[0046]段—第 5 页第[0062]段以及图 5  | 3,8,13                   |

☒ 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

“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

“&amp;” 同族专利的文件

国际检索实际完成的日期

15.01 月 2014 (15.01.2014)

国际检索报告邮寄日期

23.1 月 2014 (23.01.2014)

ISA/CN 的名称和邮寄地址:

中华人民共和国国家知识产权局

中国北京市海淀区蓟门桥西土城路 6 号 100088

传真号: (86-10)62019451

受权官员

何晓兰

电话号码: (86-10) 62085135

| C(续). 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          | US 2011/003777 A1(APPLE INC.) 17.02 月 2011 (17.02.2011)<br>说明书第 4 页第[0041]段、第 6 页第[0055]段以及图 1 | 5,10,15 |
| Y          | CN 201117218 Y (邹鹏) 17.09 月 2008 (17.09.2008)<br>说明书第 2 页第 18 行—第 3 页第 11 行以及图 1—2             | 16      |
| A          | CN 101742110 A (天津三星电子有限公司) 16.06 月 2010 (16.06.2010)<br>全文                                    | 1-16    |
| A          | JP 2002—318132 A (HITACHI LTD.) 31.10 月 2002 (31.10.2002)<br>全文                                | 1-16    |

国际检索报告  
关于同族专利的信息

国际申请号  
**PCT/CN2013/086734**

| 检索报告中引用的<br>专利文件  | 公布日期       | 同族专利               | 公布日期       |
|-------------------|------------|--------------------|------------|
| CN 102253710 A    | 23.11.2011 | 无                  |            |
| CN 1371090 A      | 25.09.2002 | 无                  |            |
| CN 102374864 A    | 14.03.2012 | US 2012/0041766 A1 | 16.02.2012 |
|                   |            | US 8412455 B2      | 02.04.2013 |
| US 2011/003777 A1 | 17.02.2011 | 无                  |            |
| CN 201117218 Y    | 17.09.2008 | 无                  |            |
| CN 101742110 A    | 16.06.2010 | 无                  |            |
| JP 2002—318132 A  | 31.10.2002 | US 2002/0156570 A1 | 24.10.2002 |
|                   |            | US 7076362 B2      | 11.07.2006 |
|                   |            | DE 60226194D D1    | 05.06.2008 |
|                   |            | DE 60226194T T2    | 14.05.2009 |
|                   |            | EP 1253569 A2      | 30.10.2002 |
|                   |            | EP 1253569B1 B1    | 23.04.2008 |

续：主题的分类

G06F 3/16 (2006.01) i

G10L 15/02 (2013.01) i