(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 104169842 B
(45)授权公告日 2017.04.05

(21)申请号 201380013687.6

(22)申请日 2013.03.05

(65)同一申请的已公布的文献号
申请公布号 CN 104169842 A

(43)申请公布日 2014.11.26

(30)优先权数据
12290086.3 2012.03.12 EP

(85)PCT国际申请进入国家阶段日
2014.09.11

(86)PCT国际申请的申请数据
PCT/EP2013/054331 2013.03.05

(87)PCT国际申请的公布数据
WO2013/135523 EN 2013.09.19

(73)专利权人 阿尔卡特朗讯公司
地址 法国布洛涅-比扬古

(72)发明人 M·法加达尔-科斯马
M·卡萨斯-桑切斯

(74)专利代理机构 北京市中咨律师事务所 11247
代理人 刘薇 杨晓光

(51)Int.Cl.
G06F 3/03(2006.01)
H04N 7/14(2006.01)

(56)对比文件
US 2011/0018963 A1,2011.07.27,
US 2010/0205667 A1,2010.08.12,
US 2007/0263077 A1,2007.11.15,
US 2008/0266385 A1,2008.10.30,
CN 101808220 A,2010.08.18,
US 2011/0018963 A1,2011.07.27,

审查员 齐银凤

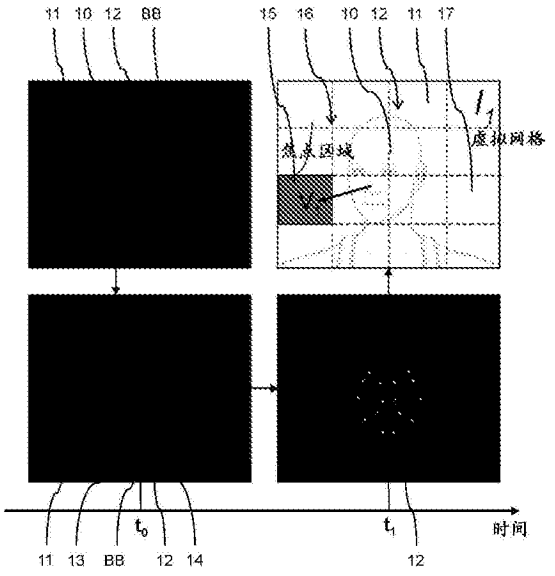
权利要求书2页 说明书9页 附图4页

(54)发明名称

用于控制视频界面的方法、用于操作视频界面的方法、面部朝向检测器以及视频会议服务器

(57)摘要

本发明涉及一种用于控制在会议情形等中使用的与用户(12)相关联的视频界面(4)的方法，包括：捕获源自所述用户(12)的视频流的帧(11)；识别所述视频帧(11)内所述用户(12)的面部(10)；检测所述视频帧(11)内所述用户(12)的所述面部(10)的朝向；以及提供指示所述面部(10)的所述朝向的控制信号。本发明还涉及一种用于操作视频界面(4)的方法，包括：上述控制方法的步骤；基于控制信号将面部(10)的朝向映射到所述视频界面(4)的焦点区域(15)；以及突出所述焦点区域(15)。本发明进一步涉及用于执行上述方法的面部朝向检测器(6)和视频会议服务器(5)。



CN 104169842 B

1.一种用于控制在会议情形中使用的与用户(12)相关联的视频界面(4)的方法,包括以下步骤:

捕获源自所述用户(12)的视频流的帧(11);

识别所述视频帧(11)内所述用户(12)的面部(10);

检测所述视频帧(11)内所述用户(12)的所述面部(10)的朝向;

提供指示所述面部(10)的所述朝向的控制信号;以及

基于控制信号将所述面部(10)的朝向映射到所述视频界面(4)的焦点区域(15)。

2.根据权利要求1所述的方法,还包括以下步骤:

执行皮肤识别以用于验证所述视频帧(11)内至少一个面部(10)的标识。

3.根据权利要求1所述的方法,其中,所述检测所述视频帧(11)内所述用户(12)的所述面部(10)的朝向的步骤包括以下步骤:

标识所述视频帧(11)内所识别的面部(10)的至少一个面部特征(14);

将所述视频帧(11)内所述至少一个面部特征(14)的当前位置与其在先前视频帧(11)中的位置进行比较;以及

根据所述视频帧(11)内所述至少一个面部特征(14)与其在先前视频帧(11)中的位置的比较,导出面部朝向。

4.根据权利要求3所述的方法,还包括以下步骤:

初始化所述所识别的面部(10)的所述面部朝向。

5.根据权利要求3所述的方法,其中,所述将所述视频帧(11)内所述至少一个面部特征(14)的当前位置与其在先前视频帧(11)中的位置进行比较的步骤包括:应用光流估计方法。

6.根据权利要求3所述的方法,其中,所述根据所述视频帧(11)内所述至少一个面部特征(14)与其在先前视频帧(11)中的位置的比较来导出面部朝向的步骤包括:基于至少一个矢量场计算朝向矢量(V_t),其中所述矢量场对于每个面部特征(14)包含一个矢量。

7.根据权利要求1至6任意一项所述的方法,还包括以下步骤:

突出所述焦点区域(15)。

8.根据权利要求7所述的方法,其中,所述将所述面部(10)的朝向映射到焦点区域(15)的步骤包括:

根据视频界面(2)提供虚拟网格(16);以及

将所述虚拟网格(16)的至少一个网孔(17)映射到所述焦点区域(15)。

9.根据权利要求7所述的方法,其中,所述突出所述焦点区域(15)的步骤包括:执行所述焦点区域(15)的放大操作。

10.根据权利要求7所述的方法,其中,所述突出所述焦点区域(15)的步骤包括:执行除了所述焦点区域(15)以外的区域的缩小操作。

11.根据权利要求7所述的方法,其中,所述突出所述焦点区域(15)的步骤包括:在所述视频界面(2)的高亮区域中显示所述焦点区域(15)的内容。

12.一种面部朝向检测器(6),包括:

用于接收视频流的视频输入(8);以及

用于提供控制信号的信令输出(8),其中所述控制信号指示所述视频流内面部(10)的

朝向；

其中，所述面部朝向检测器(6)用于执行根据权利要求1至11的任意一项的方法。

13.一种用于向用户(12)提供用户界面(4)的视频会议服务器(5)，其中，所述视频会议服务器(5)用于执行根据权利要求1至11的任意一项的方法。

14.根据权利要求13所述的视频会议服务器(5)，还包括：根据权利要求12的面部朝向检测器(6)。

用于控制视频界面的方法、用于操作视频界面的方法、面部朝向检测器以及视频会议服务器

技术领域

[0001] 本发明涉及用于控制在会议情形等中使用的与用户相关联的视频界面的方法。本发明还涉及用于操作在会议情形等中使用的与用户相关联的视频界面的方法。本发明进一步涉及面部朝向检测器，其包括用于接收视频流的视频输入和用于提供指示视频流内面部的朝向的控制信号的信令输出，其中面部朝向检测器适于执行上述方法。本发明还涉及用于向用户提供用户界面的视频会议服务器，其中，视频会议服务器适于执行上述方法。

背景技术

[0002] 对彼此地域分离的人们的通信的需求日益增长。为了便于通信和信息的交换，视频会议正变得越来越重要，以允许用户彼此交谈，看见彼此和/或交换任何类型的信息。为了提高会议结果，希望用户可以在类似会议的情形下讨论任何问题，在该情形下用户可以自然地彼此交互。

[0003] 视频会议通常基于不同用户之间的IP连接，其用于将信息从一个参加者传输到另一个。该信息通常包括能够看到和听到用户的音频/视频流，还包括将要在会议参加者之间共享的任何类型的数字文件。因此，视频会议的每个用户具有用于在本地生成用户的音频/视频流的视频摄像机，所生成的音频/视频流被提供给其它用户，并且每个用户还使用视频界面，其在本地显示在接口设备上以用于再现用户的音频/视频流和在会议中使用的任何类型的数据。

[0004] 视频会议服务器被提供以在视频会议的所有用户之间分发信息。因此，视频会议服务器将视频界面提供给用户，用户可使用任何类型的接口设备以参加视频会议，例如，用于再现音频/视频信息的屏幕和扬声器的组合。信息可例如以用户的音频/视频流的个体流(individual streams)的形式提供，或者作为包括个体流和附加文件(如果适用)的单一流(single steram)提供。

[0005] 在这种会议情形等中，用户与视频界面的交互用于改善所接收信息的表现。一个可能性是依靠连接到视频接口设备的输入设备(例如鼠标)的交互。正如已知的，鼠标可用作来自个人计算机的人机接口，以突出和操作部分视频界面，其中用户专心于或者配置视频会议本身。对于沉浸式(immersive)会议，这是不能令人满意的，因为它打破了自然交互的感觉。它要求用户随时专心于交互设备的操作，以便实现所期望的交互，并且将用户的焦点从类似会议情形的实际会议流程转移开。视频会议服务器接收来自用户的控制输入，并相应地更新他们各自的视频界面。

[0006] 另一种用于在会议情形中交互的方法基于注视控制。注视控制是指监控人类眼睛的位置，以便确定屏幕的用户聚焦的区域。注视控制依靠用户的眼睛的监控，其具有若干缺点，阻碍了该控制对于沉浸式视频会议等情形的一般使用。首先，注视控制要求高分辨率摄像机，因此并不适合于目前所用的许多普通摄像机，例如，带有摄像机的普通膝上型电脑或智能电话，其不能为注视控制提供足够的分辨率。此外，视频摄像机的视频流通常被编码以

用于通过IP连接传输。特别是在低带宽或高延迟的连接中，视频流的质量会降低，这对注视控制的准确性和性能具有负面影响。诸如眼镜或太阳眼镜的眼睛佩戴物的使用也可阻碍注视控制的使用。由于注视控制要求高质量的视频信息，因此，也要求高计算能力以处理该信息。因此，注视控制只能用提供所要求的计算能力的特定硬件来执行。

发明内容

[0007] 因此，本发明的目的是提供用于控制视频界面的方法、用于操作视频界面的方法、面部朝向检测器和视频会议服务器，其克服上述的缺点和限制。

[0008] 该目的通过独立权利要求实现。有利的实施例在从属权利要求中给出。

[0009] 具体地，提供了用于控制在会议情形等中使用的与用户相关联的视频界面的方法，其包括：捕获源自用户的视频流的帧；识别视频帧内用户的面部；检测视频帧内用户的面部的朝向；以及提供指示面部的朝向的控制信号。

[0010] 进一步地，提供了用于操作在会议情形等中使用的与用户相关联的视频界面的方法，其包括：执行如上所述的用于控制视频界面的方法；基于控制信号将面部的朝向映射到视频界面的焦点区域；以及突出焦点区域。

[0011] 此外，提供了面部朝向检测器，其包括：用于接收视频流的视频输入；以及用于提供指示视频流内面部的朝向的控制信号的信令输出；其中，面部朝向检测器适于执行上述方法。

[0012] 此外，提供了用于向用户提供用户界面的视频会议服务器，其中，视频会议服务器适于执行上述方法。

[0013] 基本思想是检测面部的朝向以用于控制和操作视频界面。面部的朝向的检测可被执行而无需强大的硬件要求，例如，生成具有特定分辨率的视频流，或者提供特定计算能力。面部的朝向的检测可基于低分辨率摄像机进行，这种摄像机是大多数膝上型电脑、智能电话或其它数据处理设备的一部分。即使提供给面部检测器的视频流是被编码的，也可适用。在会议情形等中，假定人位于摄像机的前面，以使得即使是数据低质量的视频流，也可显示足够的细节以用于面部的朝向的检测。眼睛佩戴物或其它面部佩戴物的使用仅仅部分遮盖面部，这使得面部的朝向的检测能够基于没被眼睛佩戴物或其它面部佩戴物遮盖的面部的部分。该方法适合于在云内使用或者由位于因特网中的服务器使用，因为视频流可以低数据速率提供以用于执行面部的朝向的检测。对于视频流的传输不存在高带宽要求。

[0014] 面部朝向检测器是一种设备，其可以在用户侧本地提供，例如与用于直接将视频流传递给面部朝向检测器的视频摄像机整体连接。因此，视频摄像机可提供指示面部的朝向的控制信号以及其视频流。此外，面部朝向检测器可位于远离用户的地方，例如，作为位于因特网中的网络设备。面部朝向检测器可被实现为云服务。

[0015] 面部朝向检测器要求用于接收视频流的视频输入，其可以是任何类型的合适输入。视频流可以例如直接从视频摄像机经由已知的模拟视频连接器或者从视频摄像机作为数字视频流经由IP连接而被提供为模拟或数字视频流。

[0016] 视频会议服务器产生如上所述的视频界面。视频界面的操作由用户的面部的朝向来控制。用户通常位于显示视频界面的显示器的前面，该视频界面例如可以是视频屏幕或视频屏幕的投影。视频摄像机通常位于视频界面处并面向用户，以使得用户的本地视频流

可被提供给视频会议服务器。采用该假设，控制信号可指示面部的朝向，仅仅作为例如预定义坐标系统中的一种矢量或位置。指示面部的朝向的控制信号被视频会议服务器用于提供面部的朝向到视频界面的区域的映射，其中该区域也称为焦点区域。

[0017] 焦点区域被认为是用户最感兴趣的区域，并因此被突出以便于接收在该区域中显示的信息。焦点区域可以仅仅通过显示器的点或者通过显示器的具有任何形状的区域来表示。例如，焦点区域可以是具有某一直径的圆形区域、或者方形或矩形区域。焦点区域也可以利用在视频界面上显示的视频会议的项目来定义。这种项目例如是视频会议的用户的视频流的表现、或者是由包括本地用户的视频会议的用户提供的任何类型的信息的再现。在这种情况下，面部的朝向被映射到最匹配面部的朝向的项目。

[0018] 面部检测器可例如使用HAAR分类器执行，其被应用在视频流的视频帧上。HAAR分类器对视频帧内多个面部的检测进行标记，并提供边界框作为面部的标识。优选地，具有最大尺寸的边界框被选择为用户的面部以用于进一步处理。因此，即使多个人与视频流中可见的用户在一起，也可以可靠地检测用户的面部的朝向。面部特征的标识优选使用例如Sobel或Canny的边缘算子，并应用SIFT特征检测器或“用于跟踪的好特征(good features to track)”算法。

[0019] 优选实施例还包括执行皮肤识别以用于验证视频帧内至少一个面部的标识的步骤。优选地，基于颜色的皮肤分割被应用于帧以用于执行例如由HAAR分类器识别的面部的真实性检查。因为所识别的面部的出现必须匹配皮肤颜色光谱，因此，可拒绝面部的错误出现。

[0020] 根据优选实施例，检测视频帧内用户的面部的朝向的步骤包括以下步骤：标识视频帧内所识别的面部的至少一个面部特征；将视频帧内至少一个面部特征的当前位置与其在先前视频帧中的位置进行比较；根据视频帧内至少一个面部特征与其在先前视频帧中的位置的比较，导出面部朝向。面部特征是指面部的容易跟踪的部分，例如鼻尖、下巴、嘴角或其它。将要用于本方法的面部特征的数量和种类可以根据例如视频流质量或者可用处理能力来自由选择。原则上，本方法用单个面部特征已经有效。然而，更多数量的面部特征可增加面部的朝向的检测的可靠性和准确性。为了检测面部的朝向，这些面部特征的位置在不同的视频帧之间跟踪。视频帧可以是连续的视频帧或者有延迟的视频帧。处理的视频帧越少，计算的工作量就越低，然而，连续视频帧的处理可增加面部的朝向的检测的可靠性。基于不同的面部特征的位置的差异，可导出面部朝向。在评估多个面部特征时，面部朝向可被提供为不同面部特征的朝向的变化的平均值。

[0021] 优选实施例还包括初始化所识别的面部的面部朝向的步骤。初始化可在视频会议开始时执行，或者在会议期间的任何时间执行。此外，初始化也可在视频会议期间在面部的检测丢失时执行。初始化能够实现用户的面部的可靠检测，并将用户的面部的朝向设置为预定义值，例如，指示面部朝向中心区域的空(NULL)值。

[0022] 根据优先实施例，将视频帧内至少一个面部特征的当前位置与其在先前视频帧中的位置进行比较的步骤包括：应用光流估计方法。优选地，光流估计方法是金字塔Lukas-Kanade光流估计方法。该方法容易移植到不同平台上，并进一步适合于基于GPU的执行，以使得该方法在基于云的实现中执行良好。

[0023] 根据优选实施例，根据视频帧内的至少一个面部特征与其在先前视频帧中的位置

的比较而导出面部朝向的步骤包括：基于至少一个对每个面部特征包含一个矢量的矢量场计算朝向矢量。矢量场优选地包括表示面部的旋转的旋转分量、表示面部朝向或远离摄像机移动的散度分量、以及表示平行于视频摄像机的平面的平移运动的辐射分量。优选地，这三个分量通过面部特征的光流集合的Helmholtz-Hodge分解而获得。进一步优选地，可采用Kalman滤波器以减少噪声影响。

[0024] 根据优选实施例，将面部的朝向映射到焦点区域的步骤包括：根据视频界面提供虚拟网格，以及将虚拟网格的至少一个网孔映射到焦点区域。即使没有关于由用户用于再现视频界面的显示器的知识，虚拟网格也可被提供并用于计算。焦点区域的突出优选地包括突出网格的至少一个网孔。因此，控制信号可通过标识一个网孔来指示面部的朝向。虚拟网格的网孔可根据视频会议的项目来设计。

[0025] 在优选实施例中，突出焦点区域的步骤包括：执行焦点区域的放大操作。放大或扩大可对焦点区域本身或者对焦点区域和周围区域执行。优选地，放大操作针对在视频界面上显示的视频会议的全部项目执行。

[0026] 根据优选实施例，突出焦点区域的步骤包括：执行除了焦点区域外的区域的缩小操作。根据放大操作，缩小操作可在焦点区域本身或者焦点区域和周围区域的周围执行。优选地，缩小也基于在视频界面上显示的项目。缩小可在本地例如在焦点区域周围的边界区域中或者在视频界面的除了焦点区域以外的整个剩余区域上执行。优选地，放大和缩小可被组合以用于有效地突出焦点区域。

[0027] 在优选实施例中，突出焦点区域的步骤包括在视频界面的高亮区域中显示焦点区域的内容。根据放大，焦点区域本身或者焦点区域和周围区域可在高亮区域中显示。高亮区域允许操作视频界面而无需修改其主要部分。例如，视频界面的至少一部分，例如视频界面的边界区域或者边框，可显示视频会议的所有项目，而视频界面的另一个部分，例如其中心区域，显示与焦点区域对应的项目。在可选实施例中，焦点区域的内容可被移动到高亮区域。

[0028] 根据优选实施例，视频会议服务器还包括上述面部朝向检测器。

附图说明

[0029] 现参考附图并仅以示例的方式描述根据本发明的装置和/或方法的一些实施例，其中：

[0030] 图1示出根据实施例的用于控制和操作视频界面的方法的流程图；

[0031] 图2是说明根据上述方法的检测面部的朝向的图；

[0032] 图3是说明根据上述方法的突出与焦点区域对应的视频界面的项目的图；

[0033] 图4是说明根据上述方法的基于矢量场导出面部的朝向的图；

[0034] 图5是说明根据上述方法的突出与焦点区域对应的视频界面的项目的另一个图；

[0035] 图6示出根据第一实施例的包括视频摄像机、视频会议服务器和面部朝向检测器的视频会议系统的示意图。

具体实施方式

[0036] 图6示出根据第一实施例的视频会议系统1的示意图。在该实施例中，视频会议系

统1包括视频接口设备2和数字视频摄像机3。视频接口设备2在该实施例中是LCD显示器，其再现从视频会议服务器5提供的视频界面4。视频会议系统1进一步包括面部朝向检测器6。视频接口设备2、数字视频摄像机3、视频会议服务器5和面部朝向检测器6经由IP连接7连接。在可选实施例中，面部朝向检测器6与视频会议服务器5整体地提供。

[0037] 面部朝向检测器6经由IP连接器8从数字视频摄像机3接收视频流。如以下详细描述的，面部朝向检测器6检测面部10的朝向，并经由IP连接器8将指示面部的朝向的控制信号提供给视频会议服务器5。因此，面部朝向检测器6的IP连接器8充当用于从数字视频摄像机3接收数字视频流的视频输入，以及用于提供指示在视频帧中显示的面部10的朝向的控制信号的信令输出。

[0038] 视频会议服务器5产生视频界面4，即，会议流内视频会议的再现，并经由IP连接7提供给视频接口设备2，其中示出了视频界面4的再现。

[0039] 图1示出了根据实施例的方法的流程图。方法以步骤S100开始。步骤S100包括方法的初始化，其包括初始化面部识别和在视频流中显示的面部10的朝向，如以下详细说明的。

[0040] 在初始化步骤S100，对数字视频摄像机3的视频帧11应用例如配置了Intel的OpenCV库的正面面部HAAR分类器。与时刻t_0与t_1对应的个体视频帧11在图2中示出。视频帧11显示如由会议情形中的数字视频摄像机3提供的视频会议的本地用户12，其中该本地用户12位于数字视频摄像机3的前面并面向视频接口设备2上的视频界面2。初始化包括用户12的面部10的检测和面部10的初始位置。面部检测使用正面面部HAAR分类器实施。训练普通正面面部HAAR分类器的方式要求用户12的面部10必须笔直地朝向数字视频摄像机3，以便发生检测。

[0041] 对于每个视频帧11，HAAR分类器提供面部出现的列表作为一组边界框BBi，i＝1..n，其中n表示所检测的面部出现的数量。每个BBi被表示为四元组<X,Y,W,H>，其中<X,Y>表示帧中BB中心的坐标，<W,H>表示其在图像像素中的尺寸(宽度，高度)。图2示出指示视频帧11内用户12的面部10的边界框BB。

[0042] 此外，将基于颜色的皮肤识别和分割应用于视频帧11，并通过所连接部件分析来确定皮肤碎片。然后，根据以下公式选择最大的边界框BB_{max}：

[0043] $$BB_{max}=\arg\max_{BB}\{A(BB_i)\,|\,SR_i>T_{SR}\},\ i=1..n \quad (1)$$

[0044] 其中：

[0045] -SR_i＝皮肤比(skin ratio)＝标记为皮肤的像素的数量/框区域中像素的总数；

[0046] -$A(BB_i)=BB_i.W\times BB_i.H$＝边界框面积泛函；

[0047] -T_{SR}＝专用皮肤比阈值(例如，0.8)；

[0048] -arg max＝最大化函数的参数。

[0049] 这确保了如果在场景中有多个人面向数字视频摄像机3，则只有最靠近数字视频摄像机3的人将被选择以用于进一步处理。由于来自Haar分类器的错误正面识别而导致的错误出现可以被拒绝，因为出现必须匹配皮肤颜色光谱。因此，皮肤识别提供了视频帧11内至少一个面部10的标识的验证。

[0050] 如果在视频帧11中发现BB_{max}，面部朝向矢量V_0被初始化为：

[0051] -原点＝$<BB_{max}.X,BB_{max}.Y>$；

[0052] -方向＝垂直于帧平面；

[0053] -大小＝BB_{max}.H/像素中的帧高度。

[0054] 在步骤S110，该方法继续相对于初始化而检测视频帧11中最大的面部10，如上所述的。

[0055] 在步骤S120，执行面部特征14的跟踪。因此，在图2中被标记为I_0的发生了初始面部检测的视频帧11通过边缘算子(例如，Sobel或Canny)传递以提供发生了初始面部检测的视频帧11(也称为I_0)的边缘图像E_0。边缘图像E_0包括一组边缘13。在初始面部检测后的任何时间t，当前视频帧11被称为I_t，而E_t是其对应的边缘图像。

[0056] 可被跟踪的面部特征14的特征集合F_0通过将SIFT特征检测器或者Shi和Tomasi的称为“用于跟踪的好特征”算法的算法应用于由BB_{max}定义的感兴趣区域(ROI)内的E_0来获得，如图2中所示的。

[0057] 然后，特征集合F_0在下一个边缘图像E_1中通过使用光流算法来跟踪，例如，金字塔Lukas-Kanade光流估计方法。一般地，关于边缘图像E_t的特征集合F_t通过使用光流算法估计来自集合F_{t-1}的每个面部特征14的位置来产生。

[0058] 特征集合F_t数学上表示为：

[0059] $$F_t=\{f_i \mid i=1...n_t\} \quad (2)$$

[0060] 其中，称为f_i的每个被跟踪的面部特征14被表示为四元组<x,y,x',y'>，其中，<x,y>表示集合F_{t-1}中面部特征14的先前位置，<x',y'>表示新估计的位置。考虑到$\Delta x=x'-x$和$\Delta y=y'-y$，很明显地，面部特征14可以用矢量$V^f{}_i$的形式表示，其中：

[0061] -原点＝<x,y>；

[0062] -方向＝arctg($\Delta y/\Delta x$)；

[0063] -速率＝sqrt($(\Delta x)^2+(\Delta y)^2$)。

[0064] 算法必须确保面部特征14在被跟踪一定数量的视频帧11后仍然属于用户12的面部10。这通过去除由于噪声或累积误差而造成的异常值(其是错误估计的特征)，并周期性地再生特征集合F_t以避免在去除异常值后特征集合F_t基数的减少来实现。

[0065] 异常值通过相对于帧差异$\Delta I_t=I_t-I_{t-1}$约束特征集合F_t来去除。过滤特征集合F_t中的面部特征14，以使得：

[0066] $$F_t=\{f_i \mid \Delta I_t(f_i.x',f_i.y')\neq 0\} \quad (3)$$

[0067] 特征集合F_t根据以下算法周期性地再生(在若干N_f帧后)：

[0068] -对于特征集合F_t，当t是N_f的倍数时，计算凸多边形$C(F_t)$；

[0069] -$C(F_t)$被设置为用于边缘图像E_t的ROI；

[0070] -对于在先前所考虑的ROI内的E_t再计算可被跟踪的面部特征14的集合F_t'；

[0071] -在t+1处，从F_t'开始计算跟踪。

[0072] 由于用于基于GPU执行的金字塔Lukas-Kanade流估计方法的可移植性，因此，该方法执行得非常快，并适合于服务器侧的实现。

[0073] 在步骤S130，验证所跟踪的面部特征14的集合F_t是否由于用户12的面部10移动到数字视频摄像机3的覆盖区域之外而丢失。如果所跟踪的面部特征14的集合F_t丢失，则方法返回到步骤S110，检测最大的面部10。否则，方法继续步骤S140。

[0074] 在步骤S140，根据当前分析的视频帧11更新面部朝向矢量V_t。光流算法的输出被建模为在域Ω(几乎处处都是利普希茨(Lipschitz)连续)中的矢量场u，其中可跟踪特征的

集合F_t根据下式而类似于矢量场u：

[0075] $$u=\left\{ \overline{V_i}^f \middle| f_i \in F_t \right\}$$

[0076] 在本方案中，域Ω由在其中计算了光流的边界框BB所定义的感兴趣区域给出。每个矢量场u可如下被分解(在某一组情况下，其中在该例中遇见这些情况)成3个矢量场，其也在图4中示出：

[0077] u＝d+r+h

[0078] 其中：

[0079] d＝无旋分量(即是无旋场)，

[0080] r＝无散度(纯旋转)场，

[0081] h＝谐波场(即是梯度)。

[0082] 执行由公式(3)给出的所跟踪的面部特征14的光流集合F_t的Helmholtz-Hodge分解。Helmholtz-Hodge分解产生三个分量：

[0083] -旋转分量，表示面部10的旋转；

[0084] -散度分量，表示面部10朝向数字视频摄像机3或者远离它的移动；以及

[0085] -梯度分量，表示平行于摄像机平面的纯平移运动。

[0086] Helmholtz-Hodge分解使用从存在于解决线性系统中的流体动力学中受到启发的无网格算法(meshless algorithm)来执行。

[0087] 然后，矢量场F_t的旋转、散度和谐波分量被投影为围绕以头部为中心的参考框架的旋转。这些旋转即是：

[0088] -滚动(roll)：围绕x轴旋转，

[0089] -倾斜(pitch)：围绕y轴旋转，

[0090] -偏转(yaw)：围绕z轴旋转，

[0091] 并被表示为{Δp，Δq，Δr}三元组，其存储相对先前已知的脸部朝向V_{t-1}的角度偏差。用这些值更新V_{t-1}给出了当前的头部姿态的，其也采用角度的形式表示为{p，q，r}三元组。

[0092] 使用这三个旋转分量直接作为头部姿态的指示符(即，用户12的面部10正聚焦到的点)可被改进以减少噪声的影响。噪声源自基于像素的表示的不准确和视频摄像机3的非线性。

[0093] 为了消除噪声影响，采用Kalman滤波器。直接跟踪头部姿态矢量的Kalman滤波器会涉及奇点(由于{p，q，r}三元组的角度表示)，因此，按照四元数进行公式化。四元数是R^4＝{q_1，q_2，q_3，q_4}中的矢量，表示围绕以头部为中心的参考框架的旋转。四元数和经典R^3矢量之间的转换是简单的，并对于本领域技术人员是已知的。

[0094] Kalman符号可以通过应用简化的假设而从飞行动力学中采用并改编，其中该假设是对头部的绝对位置不感兴趣，而仅仅关注它的姿态矢量。因此，离散Kalman滤波器的内部状态仅仅由四元数朝向建模。矩阵[A]和[B]从刚体的力学中采用并改变，误差矩阵[Q]、[P]和[R](过程、估计和测量误差协方差或噪声)被定义为[Q]＝σI_{4x4}，[P]仅对于t＝0是必需的，并被选择为对角线上的大值的矩阵(例如10^5)，这在数学上说明了相对于例如状态跟踪，测量在跟踪器的早期是非常重要的。矩阵[R]是：

[0095] $$\underline{R}_k = \begin{bmatrix} \sigma_\phi^2 & 0 & 0 \\ 0 & \sigma_\theta^2 & 0 \\ 0 & 0 & \sigma_\varphi^2 \end{bmatrix}_k$$

[0096] 其中，σ是实验确定的。

[0097] 在该步骤的最后部分，从Kalman滤波器获得的结果给出在三维空间中V_t矢量的方向，而面部10的边界框BB与视频帧11的大小之间的比率给出其大小$|V_t|$。这种方式得到的矢量V_t是对用户12的面部10的朝向的指示，其可以从面部检测器6利用控制信号提供给视频会议服务器5。

[0098] 在步骤S150，确定显示器4的焦点区域15。焦点区域15对应于视频界面4上用户12正聚焦的位置。因此，以面部为中心的朝向矢量V_t被投影到具有N×M个格17(也称为网孔)的虚拟网格16上。如图2中所示，虚拟网格16覆盖在视频帧11之上。计算投影是简单的，仅仅考虑矢量的X轴和Y轴分量$V^X{}_t$和$V^Y{}_t$。

[0099] 由V_t在XY平面上的投影指向的网孔17表示视频界面4上的焦点区域15。此外，$A^f{}_i$用于确定在视频界面4上显示的沉浸式通信交互式场景20中的项目18、19，如在图3和5中示出并在下面进一步说明的。

[0100] 在沉浸式视频会议中，每个用户12或参加者P_i被呈现了其交互式场景20，其被标记为S_i并可被定制。交互式场景20显示项目18、19，包括其它用户12的视频流屏幕18、称为$\{P_j, j=1..N, j<>i\}$的共享文件19和它自己的视频流屏幕18。每个视频流都经过裁剪算法，该算法将用户12的轮廓从背景中分离并在视频流中提供。该布局的目的是向每个用户12提供在同一个房间并面对其它出席者的印象。所有的处理在云中的视频会议服务器5上执行。处理流水线(PPL)保持具有每个用户12的位置的记录，每个交互式场景20(S_i)中的P_j被表示为边界框BB_j。

[0101] 通过经过上述算法监控每个用户12(P_i)的面部朝向，PPL计算焦点区域15($A^f{}_i$)，并将它覆盖在交互式场景20(S_i)之上，如图3所示。用户12必须将其面部在某一时间间隔T内对准焦点区域15的方向，以便被登记为面部10的朝向的变化。一旦面部10的朝向被登记，则PPL检查与交互式场景20中项目18、19的边界框的最大交集。

[0102] $$BB_f = \arg\max_{BB}\{\cap(BB_j) = A^f{}_i \cap BB_j \mid_j \neq i\}$$

[0103] 然后，在步骤S170，突出由BB_f表示的所聚焦的项目18、19。因此，如果所聚焦的项目18、19是视频流屏幕18，则视频流屏幕18与各自的用户12(P_i)的面部朝向矢量$|V^i|$的大小成比例地被放大。该缩放可通过缩小其它用户12的比例并通过平滑且短暂的过渡动画而将其在场景20(S_i)中重新排列来实现，如图3所示。由于PPL不断监控$|V^i|$，因此，所聚焦的视频流屏幕18的比例可以随着摄像机前的本地用户12(P_i)更靠近或者更远离视频接口设备2来调整。如果所聚焦的项目18、19是文件19，则其在交互式场景20中的位置被与$|V^i|$成比例地缩放，直到文件19占据整个交互式场景20，如图5所示的。如果在文件19已被缩放到整个场景尺寸后，$|V^i|$仍然增加(P_i移动得非常靠近视频接口设备2)，并且$|V^i| > T_{zoom}$(其中T_{zoom}是专用阈值)，则执行对文件19内容的缩放，如在图5中进一步所示的。

[0104] 在根据焦点区域15来突出项目18、19后，方法返回到步骤120。

[0105] 本发明可以体现在其它具体的装置和/或方法中。所描述的实施例在所有方面都

只是说明性的，而不是限制性的。特别地，本发明的范围由所附的权利要求指明，而不是由在此的说明书和附图指明。所有在权利要求的等同的含义和范围内的改变都被包括在其范围之内。

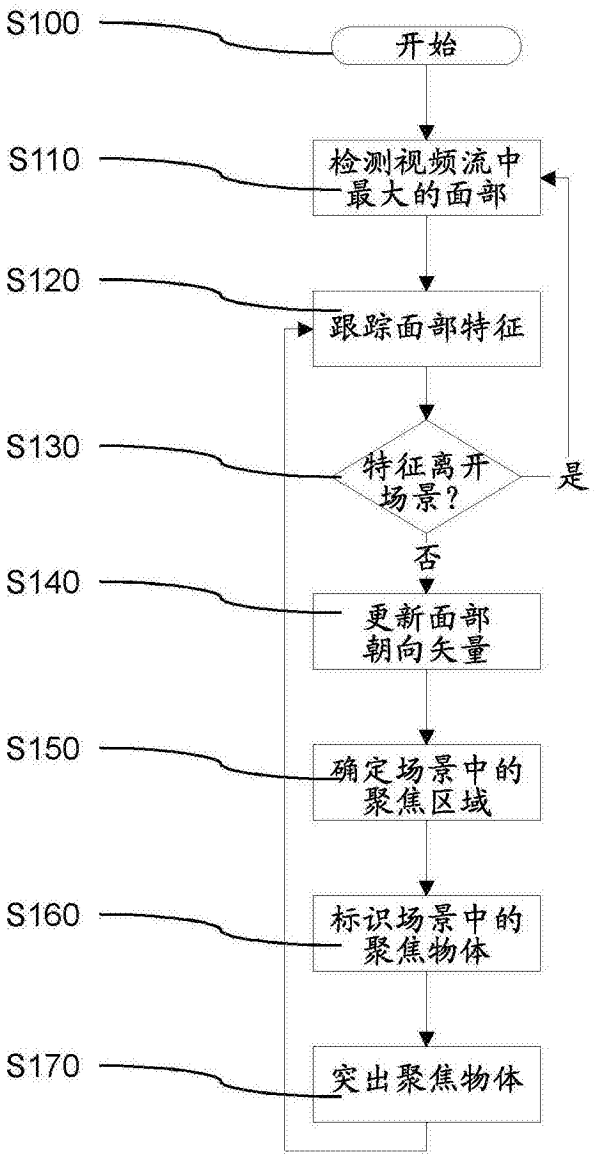

S100
开始
S110
检测视频流中最大的面部
S120
跟踪面部特征
S130
特征离开场景？
是
否
S140
更新面部朝向矢量
S150
确定场景中的聚焦区域
S160
标识场景中的聚焦物体
S170
突出聚焦物体

图1

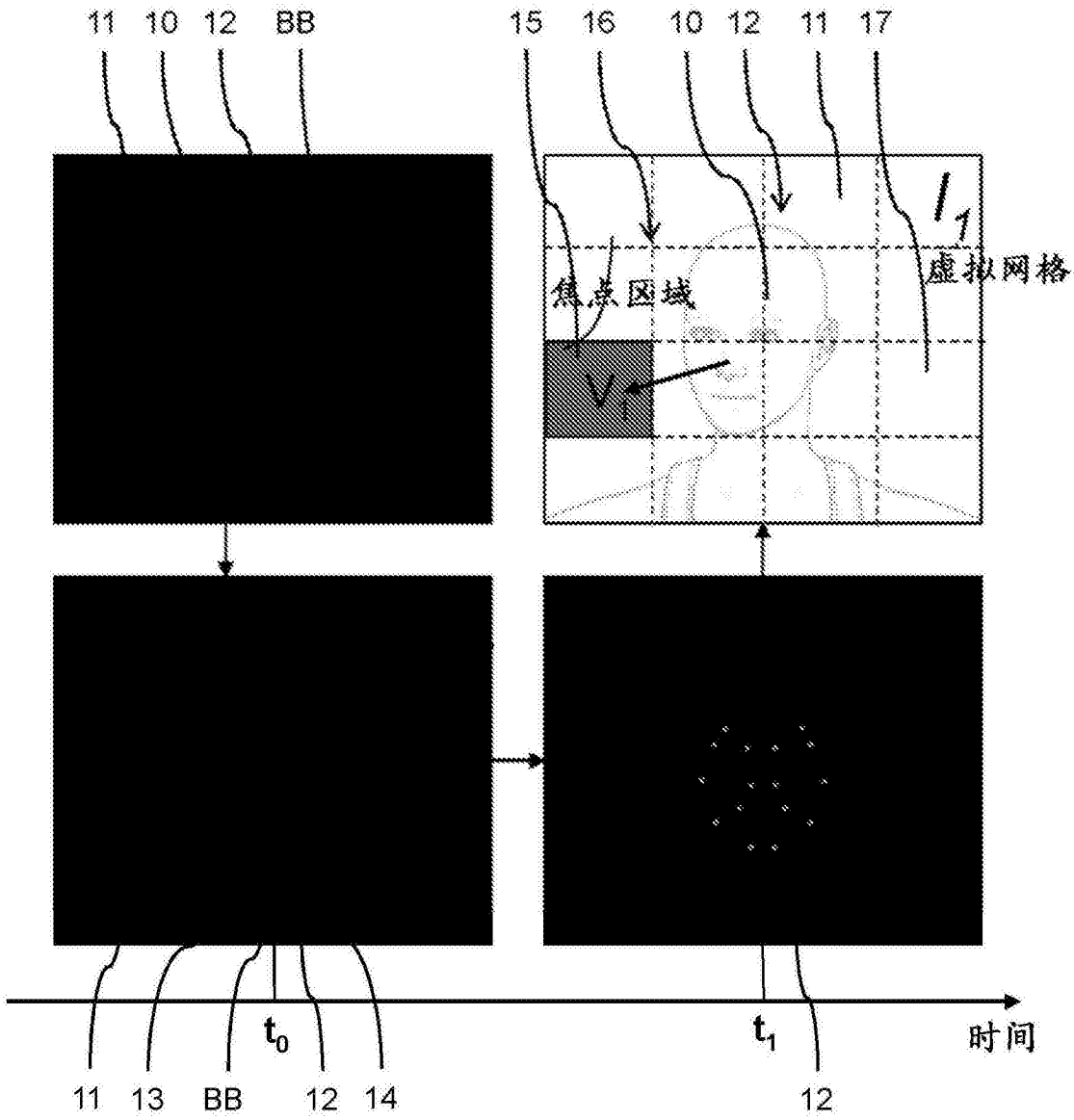

11
10
12
BB
15
16
10
12
11
17
I_1
虚拟网格
焦点区域
V_1
t_0
t_1
时间
11
13
BB
12
14
12

图2

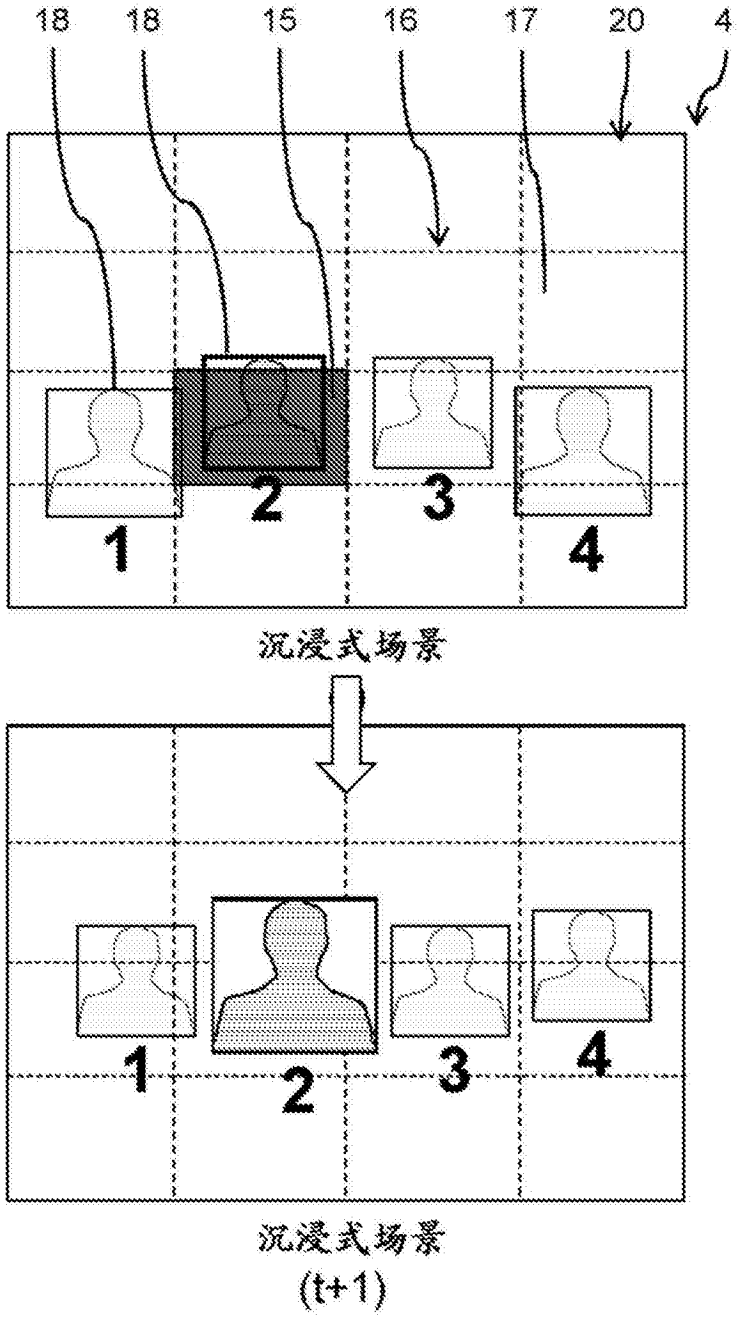


图3

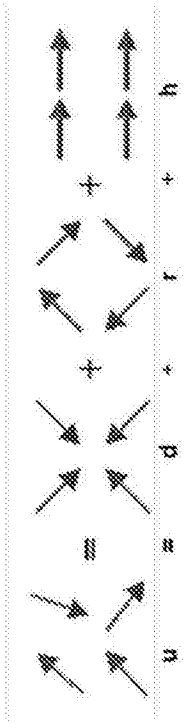


图4

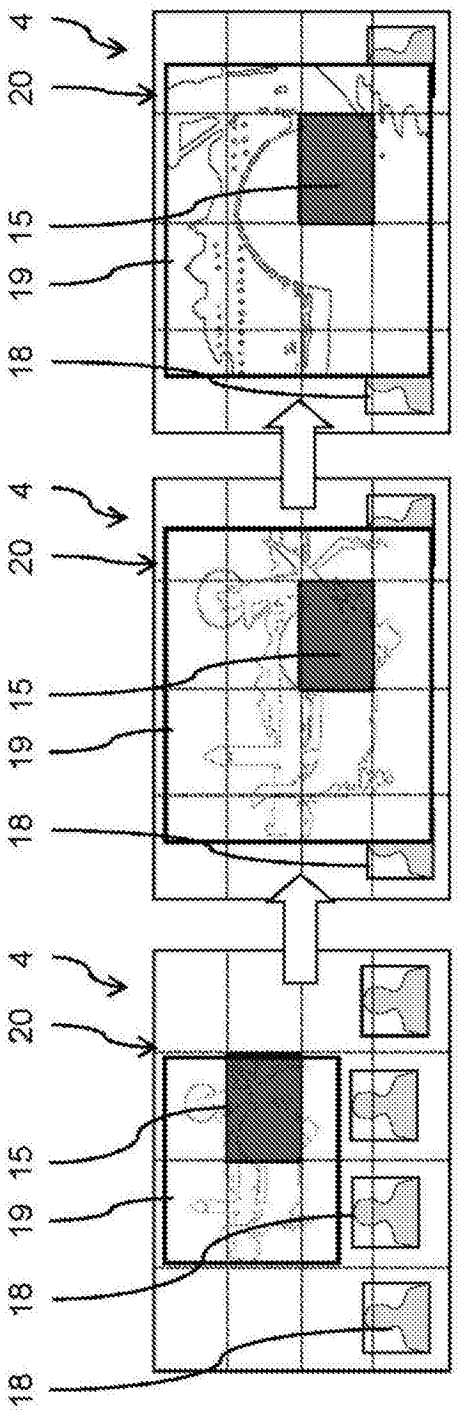


图5

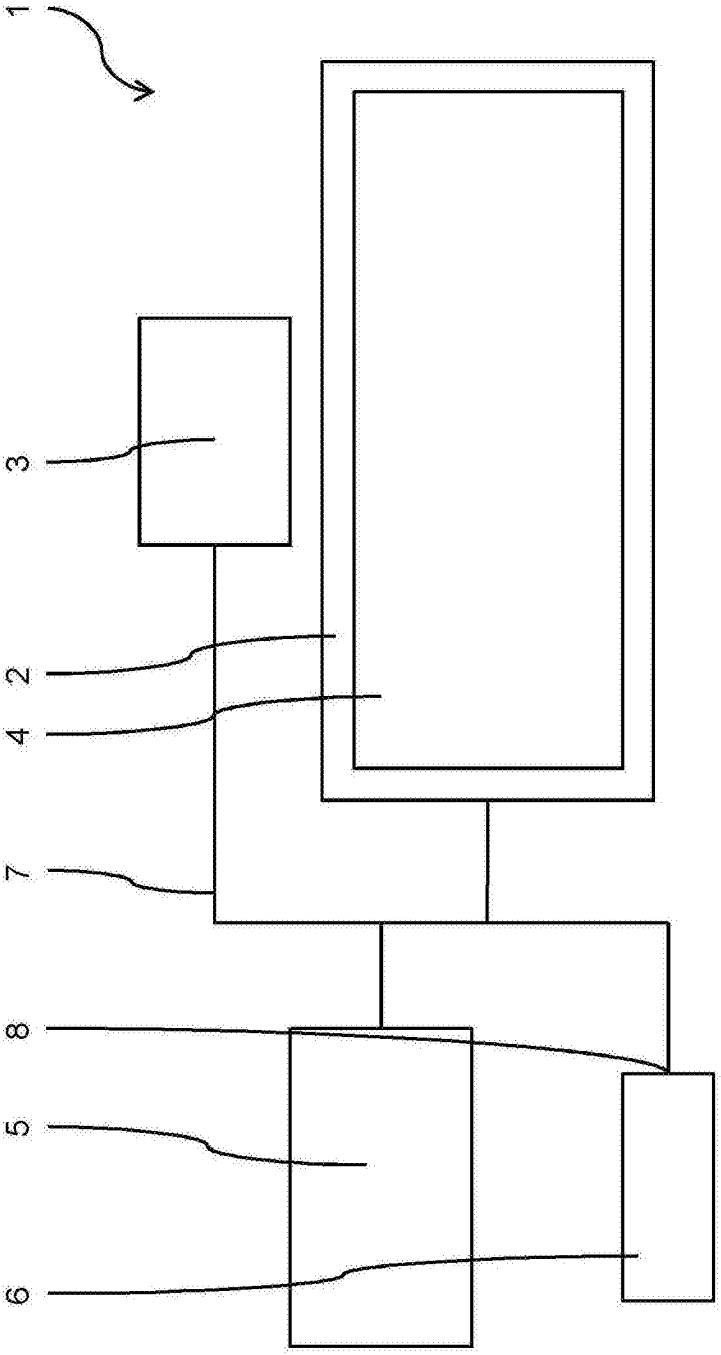


图6