(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
2009 年6 月25 日 (25.06.2009)

PCT

(10) 国际公布号
WO 2009/076853 A1

(51) 国际专利分类号:
H04N 13/00 (2006.01)

(21) 国际申请号: PCT/CN2008/073310

(22) 国际申请日: 2008 年12 月3 日 (03.12.2008)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
200710187586.7
2007 年12 月3 日 (03.12.2007) CN

(71) 申请人 (对除美国外的所有指定国): **华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人; 及

(75) 发明人/申请人 (仅对美国): **刘源 (LIU, Yuan)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **王静 (WANG, Jing)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: **北京集佳知识产权代理有限公司 (UNITALEN ATTORNEYS AT LAW)**; 中国北京市朝阳区建国门外大街22号赛特广场7层, Beijing 100004 (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, BR, BW, BY, BZ, CA, CH, 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, IS, JP, KE, KG, KM, 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, PG, PH, PL, PT, RO,

[见续页]

(54) **Title:** A THREE DIMENSIONAL VIDEO COMMUNICATION TERMINAL, SYSTEM AND METHOD

(54) **发明名称:** 立体视频通信终端、系统及方法

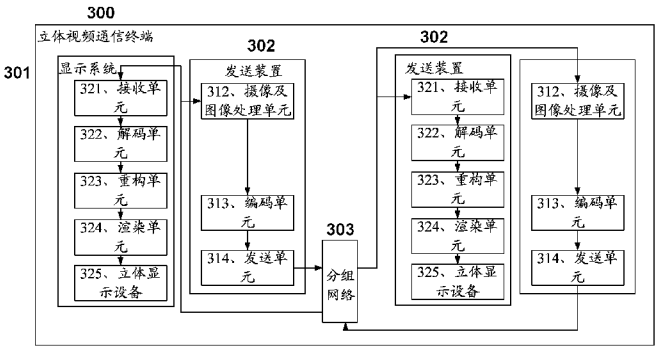


图 3 / Fig 3

300 THREE DIMENSIONAL VIDEO COMMUNICATION TERMINAL
301 DISPLAY SYSTEM
302 TRANSMITTING DEVICE
303 PACKET NETWORK
312 CAMERA AND IMAGE PROCESSING UNIT
313 ENCODING UNIT
314 TRANSMITTING UNIT
321 RECEIVING UNIT
322 DECODING UNIT
323 reconstruction UNIT
324 RENDERING UNIT
325 THREE DIMENSIONAL DISPLAY DEVICE

(57) Abstract: A three dimensional video communication terminal, system and method is disclosed. The three dimensional video communication terminal includes a transmitting device and a receiving device, in which the transmitting device includes a camera and image processing unit, an encoding unit and a transmitting unit; the receiving device includes a receiving unit, a decoding unit, a restructuring unit and a rendering unit. The three dimensional video communication system includes: a three dimensional video communication terminal, a two dimensional video communication terminal and a packet network. The three dimensional video communication method is processed in a two-way and three dimensional video communication, and it includes: shooting and acquiring video data; acquiring the depth and/or parallax information of shot object from the video data; encoding the video data and the depth and/or parallax information; packing the encoded data into the data packets according with the Real-time Transfer protocol; and transmitting the data packets via the packet network. The two-way communication of the real-time remote video stream is realized.

[见续页]

WO 2009/076853 A1

RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告。

(57) 摘要：

一种立体视频通信终端、系统以及方法。该立体视频通信终端包括发送装置和接收装置，发送装置包括：摄像及图像处理单元、编码单元及发送单元；接收装置包括：接收单元、解码单元、重构单元及渲染单元。该立体视频通信系统包括：立体视频通信终端、二维视频通信终端及分组网络。该立体视频通信方法是进行双向立体视频通信，包括：进行摄像，得到视频数据；从所视频数据中得到成像物体的深度和/或视差信息；对所述视频数据和深度和/或视差信息进行编码；将所述编码数据封装成符合实时传输协议的数据包；通过分组网络发送所述数据包。实现进行实时的远程视频流双向通信。

立体视频通信终端、系统及方法

本申请要求于 2007 年 12 月 03 日提交中国专利局、申请号为 200710187586.7、发明名称为“立体视频通信终端、系统及方法”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

本发明涉及立体视频领域，特别是涉及立体视频通信终端、系统及方法。

背景技术

立体(3D)视频技术可以提供符合立体视觉原理的具有深度信息的画面，从而能够真实地重现客观世界景象，表现出场景的纵深感、层次感和真实性，是当前视频技术发展的重要方向。

目前视频研究领域有两个主要的研究热点：双目立体视频和多视角视频（又称“Multi-View Coding，MVC”）。参阅图 1，双目立体视频的基本原理是模拟人眼成像原理，采用双摄像机独立得到左眼图像和右眼图像，并使人的左右眼分别看到左右眼通道图像，最后合成得到具有立体感的图像。多视角视频由多个（至少三个）摄像机同时拍摄得到，具有多个视频通道，不同的摄像机的拍摄角度不同，参阅图 2 是视频技术中单视角、平行多视角、汇聚多视角摄像机系统的结构示意图。在播放时这些不同角度的景物图像都会送到用户终端，如电视机显示屏上，用户在观看时可以选择不同的角度观看不同的景物画面图像。

其中，一种现有技术多视角视频技术允许用户观看动态场景，进行交互（如冻结，慢放或回退）和改变视角。应用该技术的系统采用多摄像机捕获储存的视频流，采用多视角 3D 重构和交织技术创建分层的视频帧，可以进行有效的压缩和动态场景的交互式回放。该系统还包括一个具有计算设备的渲染和接收装置，渲染程序用于根据客户端选择的视角渲染接收装置接收到的每一帧交互式视点图像。

另一种现有技术交互式多视角视频采用一种新型的视频捕获系统。该系统包括视频摄像机、控制个人计算机（PC）、服务器、网络组件和客户端，还包括音频组件用于捕获相关的音频。多个摄像机采用主－从配置方式，由一个或多个控制 PC 进行控制，从多个视点和方向进行同步采集。捕获到的视频数据

由控制 PC 进行压缩，传输到一个或多个服务器上储存。由服务器将压缩后的数据分发到终端用户或进行进一步的压缩，去除时域和空域的相关性。

在进行本发明创造过程中，发明人发现上述现有多视角视频技术至少存在以下问题：

上述现有技术多视角视频技术实现的功能单一，不能满足目前消费者的实际需求。比如现有技术一的多视角视频技术，着重于对已储存的动态场景的视频，进行交互式回放，现有技术二的多视频技术着重于将捕获的多视频数据先储存到服务器上，再分发给终端。当需要对多视角视频进行远程实时传输时，缺乏相应的系统、方法以及设备的支持；并且不支持实时双向交互式的立体视频的播放。

发明内容

本发明实施方式要解决的技术问题是提供一种立体视频通信终端、方法以及发送装置，可以进行实时的远程视频数据双向通信，多视角视频远程实时播放。

一方面，提供一种立体视频通信终端，包括：发送装置和接收装置，所述发送装置包括：摄像及图像处理单元，用于进行摄像并输出视频数据及其深度和/或视差信息；编码单元，用于进行编码，包括对所述摄像及图像处理单元输出的视频数据及其深度和/或视差信息进行编码；发送单元，用于将所述编码单元输出的编码数据封装成符合实时传输协议的数据包，并通过分组网络进行实时传输；所述接收装置包括：接收单元，用于接收数据包，包括接收发送单元发送的数据包并去除所述数据包的协议头，得到所述编码数据；解码单元，用于解码所述接收单元输出的编码数据，得到所述视频数据以及深度和/或视差信息；重构单元，用于根据所述解码单元输出所述深度和/或视差信息、以及所述解码单元输出的视频数据进行用户所处视角图像的重构，并输入到所述渲染单元；渲染单元，用于将所述重构单元输出的重构图像数据渲染到立体显示设备上。

另一方面，提供一种立体视频通信系统，包括：所述立体视频通信终端，用于进行二维或立体视频通信；二维视频通信终端，用于进行二维视频通信；分组网络，用于承载所述立体视频通信终端之间、或二维视频通信终端之间传

输的二维或立体视频数据。

再一方面，提供一种立体视频通信终端，包括：摄像及图像处理单元，进行摄像，包括进行摄像并输出视频数据及其深度和/或视差信息；编码单元，用于进行编码，包括对所述摄像及图像处理单元输出的视频数据及其深度和/或视差信息进行编码；发送单元，用于将所述编码单元输出的编码数据封装成符合实时传输协议的数据包，并通过分组网络进行实时传输。

再一方面，提供一种立体视频通信终端，包括：接收单元，用于接收数据包，包括接收发送单元发送的数据包并去除所述数据包的协议头，得到所述编码数据；解码单元，用于解码所述接收单元输出的编码数据，得到所述视频数据以及深度和/或视差信息；重构单元，用于根据所述解码单元输出所述深度和/或视差信息、以及所述解码单元输出的视频数据进行用户所处视角图像的重构，并输入到所述渲染单元；渲染单元，用于将所述重构单元输出的重构图像数据渲染到立体显示设备上。

再一方面，提供一种立体视频通信方法，包括：进行双向立体视频通信，包括：进行摄像，得到视频数据；从所视频数据中得到成像物体的深度和/或视差信息；对所述视频数据和深度和/或视差信息进行编码；将所述编码数据封装成符合实时传输协议的数据包；通过分组网络发送所述数据包。

再一方面，提供一种立体视频通信方法，包括：接收视频数据，包括：接收通过分组网络进行实时传输的视频数据包，并去除所述数据包的协议头，得到已编码立体视频编码数据；对所述已编码视频数据进行解码，得到视频数据以及相应的深度和/或视差信息；根据所述深度和/或视差信息、以及视频数据进行用户所处视角图像的重构；将所述重构的图像数据渲染到立体显示设备上。

以上技术方案可以看出，立体视频通信终端采用接收装置可以实时接收立体视频流并进行渲染，还可以采用分组网络发送立体视频数据，并进行实时传输给对方终端，用户可以远程看到实时的立体图像，实现远程立体视频通信，提升用户体验。

附图说明

图 1 是现有技术双目立体视频摄像的原理图；

图 2 是现有技术单视角、平行多视角、汇聚多视角摄像机系统的结构示意图;

图 3 是本发明立体视频通信终端第一实施方式的原理框图;

图 4 是本发明立体视频通信系统第一实施方式的原理框图;

图 5 是图 4 中发送端、接收端和分组网络两侧的原理框图;

图 6 是本发明立体视频通信系统又一实施方式的原理框图;

图 7 是本发明中在发送装置与接收装置两侧分别对视频数据进行混合编解码的流程图;

图 8 是视差、景深与用户观看距离之间的关系图;

图 9 是本发明立体视频通信方法第一实施方式发送方的流程图;

图 10 是本发明立体视频通信方法第一实施方式接收方的流程图。

具体实施方式

为使本发明的目的、技术方案、及优点更加清楚明白，以下参照附图并举实施方式，对本发明进一步详细说明。

立体视频通信终端实施例，参阅图 3，本发明实施方式提供一种双向实时，支持多视角的立体视频通信终端。通信双方的用户在使用过程中可以选择多个观看视角，并都能得到稳定的实时立体视频图像。

首先提供立体视频通信终端第一实施方式，包括发送装置、分组网络与接收装置。

所述发送装置包括:

摄像及图像处理单元 312，进行摄像，包括进行摄像并输出视频数据及其深度和/或视差信息;

编码单元 313，用于进行编码，包括对所述摄像及图像处理单元 312 输出的视频数据及其深度和/或视差信息进行编码;

发送单元 314，用于将所述编码单元 313 输出的编码数据封装成符合实时传输协议的数据包，并通过所述分组网络进行实时传输。

所述接收装置包括:

接收单元 321，用于接收数据包，包括接收所述发送单元 314 发送的数据包，并去除所述数据包的协议头，得到所述编码数据;

解码单元 322，用于解码所述接收单元 321 输出的编码数据，得到所述视频数据以及深度和/或视差信息;

重构单元 323，用于根据所述解码单元 322 输出所述深度和/或视差信息、以及所述解码单元 322 输出的视频数据进行用户所处视角图像的重构，并输入到渲染单元 324;

渲染单元 324，用于将所述解码单元 322 输出的已解码数据、或重构单元 323 输出的重构图像渲染到立体显示设备上。

为达到双向通信功能，所述网络的发送装置一侧还可以包括所述接收装置，所述网络的接收装置一侧还包括所述发送装置。

上述摄像及图像处理单元 312 可以是多视角摄像及图像处理单元; 所述发送装置与接收装置可以作为一个整体，也可以单独使用。本实施例可以实现立体视频远程实时双向通信、可适用于现场直播或娱乐表演等场景。

以上可以看出，由于采用发送单元 314 发送摄像及图像处理单元 312 摄像得到的视频数据，并经分组网络进行实时传输，在接收端可以采用接收单元实时接收所述视频数据并进行渲染，需要重构的则进行重构再进行渲染，用户可以远程看到实时的立体图像，实现远程立体视频通信，提升用户体验。

立体视频通信系统实施例，参阅图 4，是基于 H.323 协议进行组网的视频通信系统，本发明实施例中立体视频通信系统第一实施方式包括发送端、分组网络以及接收端。

所述分组网络可以实时传输视频数据。

一起参阅图 5，所述发送端包括发送装置和接收装置，所述发送装置包括:

摄像及图像处理单元 510，进行摄像，包括进行摄像并输出视频数据; 所述摄像及图像处理单元 510 可以是单视角、多视角、或单视角与多视角都支持的摄像及图像处理单元;

采集控制单元 511，用于根据指令控制所述摄像及图像处理单元 510 的操作，包括根据视频操作单元 531 发送的指令控制所述摄像及图像处理单元的操作;

同步单元 512，用于产生同步信号并输入到所述摄像及图像处理单元 510，控制其进行同步采集; 或将所述同步信号输入到所述采集控制单元 511，通知

其控制所述多摄像及图像处理单元 510 进行同步采集;

标定单元 513，用于获取所述摄像及图像处理单元 510 中摄像机的内参和外参，并输出校正指令到所述采集控制单元 511;

预处理单元 514，用于接收所述采集控制单元 511 输出的视频数据和相应的摄像机参数，并根据预处理算法对所述视频数据进行预处理;

匹配/深度提取单元 515，用于从所述预处理单元 514 输出的视频数据中得到成像物体的立体信息，并和所述视频数据一起输出给所述编码单元 516;

编码单元 516，用于进行编码，包括对所述预处理单元 514 输出的视频数据进行编码，还对所述匹配/深度提取单元 515 输出的深度和/或视差信息进行编码;

复用单元 517，用于对所述编码单元 516 输出的编码数据进行复用;

发送单元 518，用于将所述复用单元 517 输出的编码数据封装成符合实时传输协议的数据包，并通过分组网络进行实时传输。

所述接收端包括发送装置和接收装置，所述接收装置包括:

接收单元 520，用于接收数据包，包括接收所述发送单元 518 发送的数据包并去除所述数据包的协议头，得到所述编码数据;

解复用单元 521，用于对所述接收单元 520 接收的数据进行解复用;

解码单元 522，用于解码所述解复用单元 521 输出的编码数据;

重构单元 523，在所述解码单元 522 输出的已解码数据经过立体匹配技术处理时，用于对所述已解码数据进行图像重构，并输入到所述渲染单元 524;

渲染单元 524，用于将所述解码单元 522 或重构单元 523 输出的数据渲染到立体显示设备上。

在其他实施方式中，所述接收装置还可以进一步包括:

转换单元 525，用于将解码单元 522 输出的立体视频数据转换成二维视频数据;

平面显示设备 526，用于显示所述转换单元 522 输出的二维视频数据。

分组网络两侧的通信终端，用于进行通信并控制所述发送装置与立体接收装置，包括:

指令发送单元 530，用于发送指令，包括发送携带所述摄像及图像处理单

元 510 能力信息的会议发起指令; 还用于通过所述发送单元 518 向对方发送装置的采集控制单元 511 发送控制发送装置的指令，包括: 控制摄像及图像处理单元 510 内特定摄像机开关或特定视角进行摄像的指令。

视频操作单元 531，用于操作所述发送装置与接收装置，包括在收到会议确认消息后开启所述发送装置与接收装置;

多点控制单元（Multipoint Control Unit，MCU）532，连接在分组网络上，用于控制多点会议连接，包括:

能力判断单元 5320，用于在接收到所述通信终端的会议发起指令时，根据所述指令携带的能力信息判断会议双方是否具有立体摄像和立体显示能力; 在其他实施方式中，也可以将此功能集成在终端上，即不采用多点控制单元来判断会议双方/多方的能力，而是由终端自己进行判断;

会议建立单元 5321，用于在所述能力判断单元 5320 判断双方都具有立体摄像和立体显示能力时通过所述分组网络建立所述会议双方通信终端的会议连接，包括发送所述会议确认消息到双方通信终端的视频操作单元 531，以启动所述发送装置和接收装置，并将接收方通信终端的地址发送给所述发起方发送装置的发送单元 518;

转换单元 533，用于转换数据格式，包括将从一方发送装置的发送单元 518 接收到的视频数据转换成二维视频数据;

转发单元 534，用于将所述转换单元 533 输出的视频数据发送给通信对方发送装置的接收单元 520。

其中，所述多点控制系统的能力判断单元 5320 在进行能力判断的结果为会议双方有一个不具有立体显示能力时，启动所述转换单元 533 工作; 所述通信终端也同样具有能力判断功能。

其中，本实施方式可以是基于 H.323 协议进行组网的视频通信系统网络。该视频通信系统建立在分组网络上（如局域网、E1、窄带 ISDN 或宽带 ISDN 等），主要包括 H.323 网守（Gatekeeper）、H.323 网关（Gateway）、H.323 多点控制单元、普通的 2D 摄像机装置以及摄像及图像处理单元等。

（1）网守是网络中一个 H.323 实体，它为 H.323 通信终端、网关和 MCU 提供地址转换和网络接入控制。网守也可以为通信终端、网关、MCU 提供其

它服务，如带宽管理和网关定位。

（2）H.323 网关为分组网络的 H.323 通信终端、分组交换网络上的其它 ITU 终端、或者其它 H.323 网关之间提供实时双向通信。

（3）H.323 多点控制单元，如前描述，用于控制会议连接。它是网络中一个端点，为三个或更多终端及网关参加一个多点会议服务，也可以连接两个通信终端构成点对点会议，随后再扩展为多点会议。MCU 可以由两部分组成：必需的多点控制器（MC）和可选的多点处理器（MP）。其中多点控制器为多点会议提供控制功能、和通信终端进行能力协商、控制会议资源等；多点处理器在多点控制器的控制下，在多点会议中对音频、视频和/或数据流进行混合、交换等集中处理。

所述二维摄像机装置可以是 2D 视频通信终端，也可以是只具有 2D 图像采集和显示能力的视频通信终端，如可视电话，视频会议终端和个人计算机视频通信终端等。

根据上面描述，相比现有 H.323 视频通信网络，本发明实施方式中的 MCU 针对多视角立体通信系统进行改进，支持多视角立体通信系统和普通 2D 视频通信终端之间的会议控制以及立体视频流的处理。

可以理解，本发明实施例中符合实时传输的协议除了 H.323 协议之外，还包括：H.261 协议、H.263 协议、H.264 协议、会话初始协议（Session Initiation Protocol，SIP）、实时传输协议（Realtime Transport Protocol，RTP）、实时流传输协议 (Real Time Streaming Protocol，RTSP)等协议，这些并不构成对本发明的限制。

立体视频通信系统的又一实施例，参阅图 6，摄像及图像处理单元 610、采集控制单元 611、同步单元 612 和标定单元 613 构成了所述多视角立体视频通信系统的视频采集部分。其中所述摄像及图像处理单元可以是以下一种：

1）立体摄像及图像处理单元，输出包括深度和/或视差信息的视频数据；

2）分开的摄像机和匹配/深度提取单元，所述摄像机用于摄像并输出视频数据，所述匹配/深度提取单元用于从所述摄像机输出的视频数据中得到成像物体的深度和/或视差信息并输出。

对摄像及图像处理单元 610 中的摄像机进行分组，每个分组的摄像机数目

$N \geq 1$。摄像机布局采用平行多视角摄像机或环形多视角摄像机模型，从不同视点对同一场景进行拍摄。每个摄像机分组由一个采集控制单元 611 进行控制。摄像机通过摄像机数据线（CameraLink）、IEEE1394 或同轴电缆等方式连接到采集控制单元 611 进行视频流的传输，同时还通过远程控制数据线连接到指令发送单元，使用户可以远程遥控摄像机进行平移/转动/拉近/拉远等操作。整个摄像及图像处理单元 610 中，摄像机组的数目 $M \geq 1$，其根据实际应用场景的需要可自由配置，图中采用 2 组平行多视角摄像机进行视频流的传输。

同步单元 612，如前所述，用于控制多摄像机视频流的同步采集。同步单元 612 可以防止多视角摄像及图像处理单元 610 在对高速运动物体进行摄像时所获得的图像中出现图像差异问题，因为高速可导致不同视点或是同一视点左右眼的图像在同一时刻差异很大，用户看到的立体视频就会失真。同步单元 612 可以通过硬件或软件时钟产生同步信号，输出到摄像机的外同步接口对摄像机进行同步采集控制或是输出到采集控制单元 611，由采集控制单元 611 通过控制线对摄像机进行同步采集控制。同步单元 612 也可以采用一台摄像机的视频输出信号作为控制信号输入到其它摄像机进行同步采集控制。同步采集要求达到帧同步或行/场同步。

标定单元 613，如前所述，是用于完成多摄像机标定。在立体视频系统中，为了进行立体匹配和场景重建等，需要得到场景的深度或视差信息，而要求得深度或视差信息又必须知道物体某一点在世界坐标系下的坐标和其成像点坐标之间的成像关系。摄像机的内外参数是决定该成像关系的一个重要因素，而往往摄像机的内参（如图像中心、焦距，镜头畸变等）和外参又是未知、部分未知或者原则上不确定的。因此有必要通过某种方法获取摄像机的内部和外部参数，这个过程称之为摄像机标定。在立体视频的摄像机采集中，根据仿射变换原理，某一点的理想成像方程（不考虑畸变情况）可以表示为：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{R} \quad \mathbf{t}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} \qquad \mathbf{K} = \begin{bmatrix} fs & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

其中 u, v 为成像点坐标；$X_w Y_w Z_w$ 为世界坐标；s 为图像的尺度因子，为图像水平单位像素数 f_u 和垂直单位像素数 f_v 之比；f 为焦距；u_0, v_0 为图像中心坐标。R 为摄像机的旋转矩阵，t 为摄像机平移向量。其中 K 为摄像机的内参，R 和

t 为摄像机的外参。对于平行双摄像机系统，可以得到：

$$d_x(m_l,m_r)=\begin{cases}\frac{x_l}{X_l}=\frac{f}{Z}\\ \frac{x_r}{X_r}=\frac{f}{Z}\end{cases}\Rightarrow x_l-x_r=\frac{f}{Z}(X_l-X_r)=\frac{fB}{Z}$$

其中 f 为焦距，Z 为点到成像平面的距离，B 为两台摄像机光心的间距，d 为视差。可以看出，焦距 f 对深度 Z 的影响很大。此外内参中的图像中心和畸变系数等也会对深度和/或视差计算有影响，需要利用这些参数进行图像校正。

本实施方式可采用多种方法对摄像机进行标定，比如采用传统标定法和自标定法等。传统标定法包括 70 年代初提出的直接线性变换（DLT）标定法和基于径向排列约束（RAC）的标定法等。其基本方法是建立摄像机成像模型线形方程组，并测出场景中一组点的世界坐标和其在成像平面上的对应坐标，然后将这些坐标值代入该线形方程组中求出内参和外参。自标定是指不需要标定块，仅仅通过图像点之间的对应关系就可以对摄像机进行标定的过程。自标定依据的是多幅图像中成像点间存在的特殊约束关系（如极线约束），因此可以不需要场景的结构信息。自标定法的优点是灵活，方便。

本发明实施方式中标定单元 613 的功能是完成多摄像机标定，求出每个摄像机的内参和外参，针对不同的应用场景采用不同的标定算法。如针对视频会议场景，所述标定单元 613 可采用改进后的传统标定法进行标定，能够简化传统标定法的复杂的处理过程，但相比自标定法能够提高精度和减少标定时间。其基本思想是提供或找到一种和拍摄场景相融合，并且能够稳定存在的物品作为参照物，如视频会议场景中用户的名牌，或是场景中杯子等物品。这些物品应该可以提供物理几何尺寸，并且有丰富的特征可供提取。如名牌的边缘和上面的文字或图案，杯子的同心圆特征等，并采用相应的算法进行标定。如采用平面标定法进行标定的方法如下：

（1）提供已知物理尺寸的平面标定参照物；

（2）从不同角度拍摄得到平面标定参照物的图像；

（3）自动匹配并检测出平面标定参照物图像上的特征点，如文字和图案的特征点等；

（4）根据平面标定算法求出摄像机的内参和外参；

（5）求出畸变系数，并进行优化。

为避免出现不同摄像机的参数相差较大的情况，如摄像机的焦距和外参，在本发明更多实施方式中，可以将求出的不同摄像机内参和外参作为反馈信息提供给采集控制单元，采集控制单元根据当前参数的差异对摄像机进行调节，通过迭代过程使不同摄像机的参数差异达到可接受的水平。

采集控制单元 611，如前所述，用于控制一个摄像机组进行视频图像的采集并输出。摄像机组的数目可以根据场景和需求进行配置，当摄像机数为 1 时，采集控制单元输出 2D 视频流，当摄像机数为 2 时，采集控制单元输出双目立体视频流，当摄像机数大于 2 时，输出为多视角视频流。对于模拟摄像机，采集控制单元将模拟图像信号转换为数字视频图像。图像以帧的形式保存在采集控制单元的缓存中。此外，采集控制单元 611 还将采集到的图像提供给标定单元 613 进行摄像机标定，标定单元 613 将得到摄像机内参和外参返回给采集控制单元 611。采集控制单元 611 根据这些参数建立起视频流和所属采集摄像机属性的一一对应的关系。这些属性包括摄像机唯一的编号、摄像机内参和外参、每帧的采集时间戳等，并将摄像机属性和视频流按照一定格式进行输出。除了上述功能，采集控制单元 611 还提供对摄像机的控制功能和图像采集的同步功能：采集控制单元 611 可以根据摄像机标定的参数，通过摄像机的遥控接口对摄像机进行平移/转动/拉近/拉远等操作；采集控制单元 611 也可以通过摄像机的同步接口向摄像机提供同步时钟信号用以控制同步采集。此外，采集控制单元 611 也可以接受输入控制单元 620 的控制，如根据用户选择的视点信息关闭不需要的摄像机的视频采集。

预处理单元 614，如前所述，对采集得到的视频数据进行预处理，具体是从采集控制单元 611 处接收采集的图像缓存和相应的摄像机参数，根据预处理算法对缓存的图像进行处理。预处理的主要内容包括：去除图像的噪声；消除不同摄像机带来的图像差异，如对不同摄像机设置导致的图像亮度和色度差异进行调节；根据摄像机内参中的畸变系数对图像进行校正，如径向畸变校正；和/或对于某些依赖扫描线匹配的立体匹配算法（如 Dynamic Programming 动态规划法）进行扫描线对齐操作等。预处理后的图像可以消除大部分采集时带来的图像噪声和由于摄像机差异导致的不希望的图像间的不一致性，有利于后续的立体匹配和深度/视差提取。

匹配/深度提取单元 615，如前所述，用于从所述预处理单元 614 输出的视频数据中得到成像物体的三维信息，并和所述视频数据一起输出给视频编码/解码单元 616；立体图像匹配是立体视频中的一项关键技术。立体视频的重构需要得到成像物体的三维信息，而其中关键的深度信息必须从多幅图像中获取。获取深度信息的首要工作就是需要找到场景中某一点在多幅图像中对应的成像点，然后再根据该点在多幅图像中坐标求出其在空间中的坐标，从而得到该点的深度信息。找到场景中某一点在不同图像中对应成像点的过程由图像匹配完成。

本发明实施方式可采用的立体匹配技术主要包括：基于窗口的匹配、基于特征的匹配或动态规划法等。其中基于窗口的匹配和动态规划法都采用了基于灰度的匹配算法。基于灰度的算法的基本思想是将图像分割成小的子区域，以其灰度值作为模版在其它图像中找到和其最相似灰度值分布的子区域，如果两个子区域满足相似性要求，可以认为子区域中的点是匹配的。在匹配过程中，可以使用相关函数衡量两个区域的相似性。基于灰度的匹配一般都得到图像的密集的深度图。基于特征的匹配没有直接利用图像的灰度，而是利用由图像灰度信息导出的特征进行匹配，相比利用简单的亮度和灰度变化信息更加稳定。匹配特征可以认为是潜在的能够描述场景 3D 结构重要特征，如边缘和边缘的交点（角点）。基于特征的匹配一般先得到稀疏的深度信息图，然后利用内插值等方法得到图像的密集深度信息图。

匹配/深度提取单元 615 的主要功能是对两个相邻摄像机采集的视频图像进行匹配，并计算得到视差/深度信息。匹配/深度提取单元 615 对两个邻近摄像机拍摄图像的最大视差（Parallax）进行了限制，如果超过该最大视差，匹配算法的效率会很差，无法得到较高精度的视差/深度信息。该最大视差可由系统预先进行配置。本发明实施方式中，匹配/深度提取单元 615 所采用的匹配算法可以根据实际情况改变，从多种匹配算法（如窗口匹配，动态规划法等）中进行选择，根据应用场景进行配置。完成了匹配操作之后，匹配/深度提取单元 615 根据得到的图像视差和摄像机的参数求出场景的深度信息。下面以基于灰度的窗口匹配算法为例进行说明：

设 $f_L(x,y)$ 和 $f_R(x,y)$ 作为左右摄像机采集的两幅图像，(x_L,y_L) 为 $f_L(x,y)$ 中的

一点。以(x_L,y_L)为中心构成一个模版 T，大小为 m×n。可以在$f_R(x,y)$中平移该模版，设在水平位置平移Δx，在垂直位置平移Δy，其在$f_R(x,y)$中覆盖的第 k 个区域为S_k，则S_k和 T 的相关性可以用相关函数衡量:

$$D(S_k,T)=\sum_{i=1}^{m}\sum_{j=1}^{n}[S_k(i,j)-T(i,j)]^2=\sum_{i=1}^{m}\sum_{j=1}^{n}[S_k(i,j)]^2-2\sum_{i=1}^{m}\sum_{j=1}^{n}S_k(i,j)T(i,j)+\sum_{i=1}^{m}\sum_{j=1}^{n}[T(i,j)]^2$$

当$D(S_k,T)$达到最小时，可以认为达到最佳匹配。如果S_k和 T 相同，则$D(S_k,T)=0$。

在上式中，$\sum_{i=1}^{m}\sum_{j=1}^{n}[T(i,j)]^2$代表了模版 T 的能量，为常数，$\sum_{i=1}^{m}\sum_{j=1}^{n}[S_k(i,j)]^2$为$S_k$区域的能量，其随着模版 T 的变化而变化。如果 T 的变化范围较小，$\sum_{i=1}^{m}\sum_{j=1}^{n}[S_k(i,j)]^2$也近似为常量。为了使$D(S_k,T)$最小，则$\sum_{i=1}^{m}\sum_{j=1}^{n}S_k(i,j)T(i,j)$应该最大。此处采用归一化交叉相关算法(Normalized Cross Correlation，NCC)消除亮度差异引起的误匹配，相关函数可表示为:

$$C(\Delta x,\Delta y)=\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}|S_k(i,j)-E(S_k)\|T(i,j)-E(T)|}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}[S_k(i,j)-E(S_k)]^2}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}[T(i,j)-E(T)]^2}}$$

其中$E(S_k)$和$E(T)$分别为S_k和 T 的平均灰度值。当$C(\Delta x,\Delta y)$最大时，$D(S_k,T)$最小，可以认为(x_L,y_L)和点$(x_L+\Delta x,y_L+\Delta y)$匹配。$\Delta x,\Delta y$分别为两幅图像之间的水平视差和垂直视差。对于上面所述的平行摄像机系统，垂直视差近似为 0，水平视差可以表示为：$\Delta x=\frac{fB}{Z}$，从而可以求出场景中某点的深度信息：$Z=\frac{fB}{\Delta x}$。

在其他实施方式中，匹配/深度提取单元 615 可以对匹配算法进行优化（如采用并行计算），以保证系统的实时性。

视频编码/解码单元 616，如前所述，对视频数据进行编码/解码，包括视频编码单元和视频解码单元。在本发明实施方式中，立体视频编码主要可以分为两类：基于块的编码和基于对象的编码。在立体图像的编码中，除了帧内预测和帧间预测消除空域和时域上的数据冗余度外，还可以消除多通道图像之间的空域数据冗余性。比如，可以通过视差估计与补偿来消除多通道图像间的空域冗余度。视差估计补偿的核心是找到两幅（或多幅）图像间的相关性，其与运动估计补偿是类似的。

本发明实施方式所述的视频编解码单元对多视角视频数据按以下两种方

式中的一种进行编解码:

1）对不同视点间图像的视差小于等于所设定的最大视差的情况，按一帧图像+视差/深度值+部分残差的混合方式进行编码（其中视差/深度描述采用MPEG Part 3: Auxiliary video data representation 标准），图 7 显示了对于双目立体视频采用混合编码方案的基本流程实例，图中编码端获得左右图像和图像的视差/深度信息，可以采用传统方式对左图像、视差/深度信息进行编码，还可以参考左图像的编码对右图像进行预测编码，将编码后的数据传输到解码端；解码端对左图像数据、视差/深度信息及右图像残差数据解码，并合成显示立体图像。

2）对于不同视点间图像的视差大于所设定的最大视差的情况，按传统方式（如 H.263、H.264 编解码标准）对视频流进行独立编码。混合编解码方案充分利用了相邻图像间的相关性，压缩效率更高，能减少更多的相邻图像间存在的时域和空域数据冗余。此外，采用视差/深度编码有利于进行图像的重构。对于图像中有遮挡，无法提取视差/深度的部分，采用残差编码，使重构图像的质量更好。如果不同视点间图像的视差过大，这时采用运动估计和补偿等传统方法对不同视点的视频流进行独立编码，如 MPEG 标准组织的 MVC 编码标准。此外，本发明所述的编解码单元还支持可伸缩视频编码（SVC）编码标准，使系统对不同网络情况的适应性更好。

此外，视频编解码单元还接收来自输入控制单元 620 的反向信道的输入，根据用户的信息对编解码进行控制。基本的控制包括下面两方面:

（1）根据用户选择的视点，找到相应的视频流进行编码，对于用户没有观看的视点的视频流不进行编码，通过这种方式可以有效地节省视频编解码单元处理能力;

（2）根据用户终端的显示能力对相应的视频流进行编解码，如对于只具有 2D 显示能力的终端，则编码并发送 1 路 2D 视频流。通过这种方式可以提高所述多视角立体视频通信系统和普通视频通信系统的兼容性，并减少无用的数据的传输。

复用/解复用单元 617，如前所述，包括复用单元和解复用单元，复用单元接收来自视频编解码单元的已编码视频流，按帧/场的方式对多路视频流进行

复用，如按场方式进行复用，可以将一个视频流编码为奇场，另一个视频流编码为偶场，奇偶场作为一帧进行传输；解复用单元接收来自接收单元的分组数据，进行解复用，还原多路已编码视频流。

发送/接收单元 618，如前所述，包括发送单元和接收单元，可以称为网络传输单元，发送方的发送单元接收复用单元的复用数据流，对数据流进行分组处理，封装成符合 RTP 等协议的数据包，通过网络接口（如以太网接口、ISDN 接口等）向外发送。此外，发送方的发送单元还接收音频编码/解码单元 621 提供的编码音频数据流、系统控制单元 622 发送的信令数据流和用户数据单元 623 提供的用户数据（如传输的文件数据），对这些数据进行分组并通过网络接口发送到接收端。接收端的接收单元接收到发送端的分组数据后，去除协议头保留用户有效数据，并按数据类型分别发送到解复用单元、音频解码单元、系统控制单元 622 和用户数据单元 623。另外，对每一种媒体类型，完成适当的逻辑成帧、顺序编号、差错检测和差错纠正。

重构单元 630，如前所述，用于对所述解码单元输出的已解码数据进行图像重构，并输入到所述渲染单元。它的功能主要有两个：

（1）解决用户如何观看到没有放置摄像机视点的视频图像的问题，因为摄像机的数目是有限的，可能没有覆盖所有的视点，而用户有可能需要观看没有放置摄像机视点处的场景。重构单元 630 可以从输入控制单元 620 得到用户观看的视点信息，如果用户选择的视点位于摄像机的拍摄视点上，则重构单元 630 不进行重构；如果用户选择的视点位于两个邻近的摄像机组或同组的邻近摄像机之间（没有摄像机的虚拟视角），则重构单元 630 根据相邻的摄像机拍摄的图像重构用户选择视点处的图像。通过某个摄像机拍摄视点的视差/深度信息，以及相邻摄像机的位置参数信息，根据投影方程确定该场景的点在某个虚拟视角中的成像点坐标，从而可以重构出该虚拟视角的视频图像。

（2）解决用户通过自动立体显示器观看立体图像因为位置移动带来的视差变化，从而导致所看到的立体图像发生变化的问题。自动立体显示器可以使用户在不带眼镜的条件下也能看到立体图像，但此时用户离自动立体显示器的距离是可以改变的，导致图像的视差发生变化。

这里，有必要说明视差、景深和用户观看距离之间的关系。参阅图 8，图

中显示了平行摄像机系统下图像视差 p 和物体景深 z_p 以及用户离显示器距离 D 的关系，通过简单的几何关系可以得到:

$$\begin{cases} \frac{x_L}{D}=\frac{x_p}{D-z_p} \\ \frac{x_R-x_B}{D}=\frac{x_p-x_B}{D-z_p} \end{cases} \Rightarrow \frac{x_L-x_R+x_B}{D}=\frac{x_B}{D-z_p} \Rightarrow |x_L-x_R|=x_B\left(1-\frac{D}{D-z_p}\right)=x_B\left(\frac{1}{\frac{z_p}{D}-1}+1\right)=p$$

从上式可以看出，图像的视差 p 依赖于用户到显示器的距离 D。立体视频接收端收到的立体视频图像一般只具有固定的视差，可以作为一个参考视差 p_{ref}，当 D 发生改变时，重构单元对视差 p_{ref} 做相应的调整，生成新的视差 p'，并根据新视差重新生成另一幅图像。这样可以保证用户和显示表面的距离发生改变时能看到合适的图像。用户到显示表面的距离可以通过摄像机算出深度图自动检测，或由用户通过输入控制单元 620 进行手工控制。

输入控制单元 620，主要接收通信终端的输入，并反馈给采集控制单元 611、编码单元和重构单元 630，用于控制多视频流的编码和重构。输入控制单元 620 包括的信息主要有观看视点、显示和用户的距离信息等。输入控制单元 620 的信息可以由最终用户通过 GUI 界面或遥控设备进行输入，如观看视点、距离信息和显示方式；也可由终端自己检测，如终端的显示能力信息。

渲染单元 631 和显示设备，如前所述，渲染单元 631 接收重构单元 630 提供的视频数据流，将视频图像渲染到显示设备上。本发明所述的多视角立体视频通信系统支持多种显示终端，包括普通的 2D 视频显示设备、自动立体显示设备、立体眼镜和全息显示设备等。

此外，在其他实施方式中，还可以包括其它系统单元，比如：音频编码/解码单元 621、用户数据单元 623 和系统控制单元 622，功能如下:

音频编码/解码单元 621（G.711、G.729 等）: 对来自通信终端的麦克风的音频信号进行编码发送，另外，对从接收单元接收到的音频码进行解码并输出到扬声器。

用户数据单元 623: 支持远程信息处理应用，如电子白板、静态图像传输、文件交换、数据库访问、音频图形会议等。

系统控制单元 622: 为终端的正确的操作提供信令。它提供呼叫控制、能力交换、命令和指示的信令、以及消息。

在该网络结构中，当一方发起视频通信会话时，首先通过多点控制单元或自身与对端进行能力协商。如果双方都是多视角立体视频通信系统，则用户双方都可以从不同视点看到实时的立体视频，如果一方是普通的 2D 视频通信终端，由于不满足立体视频通信的条件，在多点控制单元控制下，双方用户只能以 2D 方式进行视频通信。

在进行多视角视频通信时，多视角立体通信系统有两种显示模式:

（1）在单视频图像显示模式下，接收端用户可以通过指令发送单元的 GUI 界面或遥控器等方式对观看视点进行选择，然后由通信终端将观看视点的信息通过信令发送到对端，对端的采集控制单元 611 接收到指令后，对摄像及图像处理单元 610 进行操作，或在接收的视频数据中选择相应视点的视频流输出，经选择的视频流经编码并传输回接收端的显示设备。用户看到的视频画面可以是一个立体图像（左右两幅图像，由多视角视频摄像及图像处理单元中的两台摄像机采集图像），也可以是一个 2D 图像;

（2）在多视频图像显示模式下，接收端用户可以在发送端的多视角视频摄像及图像处理单元工作时时，同时从不同的视点观看对方的场景，系统同时显示多个图像。

值得说明的是，前述本发明立体视频通信终端第二实施方式中的各单元可以集成在一个处理模块中，比如将采集控制单元 611、预处理单元 614、匹配/深度提取单元 615、视频编码/解码单元 616、复用/解复用单元 617、发送/接收单元 618 集成在一个处理模块中；同理，前述本发明立体视频通信终端、多视角视频装置其他实施方式的各单元也可以集成在一个处理模块中；或者，前述各实施方式各单元中的任何两个或两个以上都可以集成在一个处理模块中。

还值得说明的是，本发明实施方式中的各单元既可以采用硬件的形式实现，可软件实现的部分也可以采用软件功能模块的形式实现。相应地，本发明电话网关实施方式既可以作为独立的产品使用，可软件实现的部分也可以存储在一个计算机可读取存储介质中进行使用。

立体视频通信方法的实施例，参阅图 9 和图 10，本发明还提供立体视频通信方法第一实施方式，分别包括发送方和接收方的流程，包括以下步骤:

进行双向立体视频通信，包括发送视频数据与接收视频数据。

其中，参阅图 9，发送视频数据包括:

步骤 802：进行摄像，得到视频数据;

步骤 806：从所视频数据中得到成像物体的深度和/或视差信息;

步骤 807：对所述视频数据和深度和/或视差信息进行编码;

步骤 808：对所述已编码视频数据进行复用;

步骤 809：将所述编码数据封装成符合实时传输协议的数据包，通过分组网络发送所述数据包。

在其他实施方式中，所述步骤：进行摄像得到视频数据是进行多视角摄像得到多视角视频数据。

在所述步骤 807：对所述视频流进行编码之前包括:

步骤 801：对所述多视角摄像进行图像同步处理。

在步骤 802：对同步后的摄像得到的图像进行采集之后包括:

步骤 803：对采集后的多幅图像进行摄像机标定并返回摄像机参数到图像采集处理；即获取摄像机的内参和外参，并根据所述内参和外参校正所述摄像操作;

步骤 804：对采集得到的图像进行预处理;

步骤 805：判断是否满足视差限制条件;

步骤 806：满足视差限制条件时，进行立体匹配、视差/深度提取，即提取成像物体的立体信息，然后对所述视频流进行编码;

步骤 807：不满足视差限制条件时，直接对视频流进行编码。

在其他实施方式中，在发送所述封装数据前，可以进一步包括以下步骤:

步骤 808：对已编码的视频流进行复用。

所述步骤：进行双向立体视频通信还包括：发送携带所述摄像及图像处理单元能力信息的会议发起指令。

在所述步骤 809：通过分组网络发送所述数据包之后进一步包括:

根据接收到的所述会议发起指令及其携带的能力信息，判断会议双方是否具有立体摄像和立体显示能力;

在判断双方都具有立体摄像和立体显示能力时，通过所述分组网络建立所述会议双方通信终端的会议连接，启动通信双方的摄像及图像处理单元和接收

装置。

在判断双方中的一方不具有摄像能力时，进一步包括:

将发送方的视频数据转换成二维视频数据并发送给接收方。

参阅图 10，接收视频数据包括:

步骤 901: 接收通过分组网络进行实时传输的视频数据包，并去除所述数据包的协议头，得到已编码立体视频编码数据;

步骤 903: 对所述已编码视频数据进行解码，得到视频数据以及相应的深度和/或视差信息;

步骤 905: 根据所述深度和/或视差信息、以及视频数据进行用户所处视角图像的重构;

步骤 906、907: 将所述重构的图像数据渲染到立体显示设备上。

在其他实施方式中，

在去除所述数据包的协议头之后，解码之前，进一步包括:

步骤 902: 判断所述数据包是否是复用的视频数据，在判断为是时对所述复用数据包进行解复用。

在其他实施方式中，在步骤: 渲染到立体显示设备上之前，进一步包括:

步骤 904: 判断所述已解码数据是否需要进行图像重构;

在判断为需要进行重构时，进行步骤 905，对所述已解码数据进行图像重构; 在判断为不需要进行重构时，进行步骤 906、907，将解码后的数据渲染到立体显示设备上。

此外，对所述已编码视频数据进行解码之后，可以进一步包括: 判断本端显示设备是否具有立体显示能力，在判断为否时，将解码后的立体视频数据转换成二维视频数据并向平面显示设备发送。

综上所述的立体视频通信终端、系统及方法，本发明至少可以产生如下技术效果: 可以实现立体视频远程实时双向通信、可适用于现场直播或娱乐表演等场景; 可以实现双向、实时的多视角立体视频通信，适用于家庭通信、商务会议等场景; 充分利用网络资源，并且用户在使用多视角视频通信过程中可以选择多个观看视角，完全区别于现有技术视频通信方式，让用户得到亲临现场的感觉，提升用户体验。

本领域普通技术人员可以理解实现上述立体视频通信方法实施方式中的全部或部分步骤是可以通过程序来指令相关的硬件来完成，所述的程序可以存储于计算机可读取存储介质中，该程序在执行时，可以包括前述本发明立体视频通信方法各个实施方式的内容。这里所称得的存储介质，如：ROM/RAM、磁碟、光盘等。

以上对本发明所提供的一种立体视频通信终端、系统及方法通过具体实施例进行了详细介绍，以上实施例的说明只是用于帮助理解本发明的方法及其思想；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本发明的限制。

权利要求

1、一种立体 3D 视频通信终端，其特征在于，包括发送装置和接收装置，所述发送装置包括:

摄像及图像处理单元，用于进行摄像并输出视频数据及其深度和/或视差信息;

编码单元，用于进行编码，包括对所述摄像及图像处理单元输出的视频数据及其深度和/或视差信息进行编码;

发送单元，用于将所述编码单元输出的编码数据封装成符合实时传输协议的数据包，并通过分组网络进行实时传输;

所述接收装置包括:

接收单元，用于接收数据包，包括接收对端发送单元发送的数据包并去除所述数据包的协议头，得到所述编码数据;

解码单元，用于解码所述接收单元输出的编码数据，得到所述视频数据以及深度和/或视差信息;

重构单元，用于根据所述解码单元输出所述深度和/或视差信息、以及所述解码单元输出的视频数据进行用户所处视角图像的重构，并输入到所述渲染单元;

渲染单元，用于将所述重构单元输出的重构图像数据渲染到立体显示设备上。

2、根据权利要求 1 所述的立体视频通信终端，其特征在于，所述摄像及图像处理单元是单视角、多视角、或单视角与多视角都支持的摄像及图像处理单元。

3、根据权利要求 1 所述的立体视频通信终端，其特征在于，进一步包括:

分组网络两侧的通信终端，用于进行通信并控制所述发送装置与接收装置，包括:

指令发送单元，用于发送指令，包括发送携带所述摄像及图像处理单元能力信息的会议发起指令;

视频操作单元，用于操作所述发送装置与接收装置，包括在收到会议确认消息后开启所述发送装置与接收装置。

4、根据权利要求 3 所述的立体视频通信终端，其特征在于，所述发送装置进一步包括:

采集控制单元，用于根据指令控制所述摄像及图像处理单元的操作，包括根据所述视频操作单元发送的指令控制所述摄像及图像处理单元的操作。

5、根据权利要求 1 所述的立体视频通信终端，其特征在于，所述通信终端的指令发送单元进一步用于通过所述发送单元向对方发送控制发送装置的指令。

6、根据权利要求 5 所述的立体视频通信终端，其特征在于，所述控制发送装置的指令包括: 控制摄像及图像处理单元内特定摄像机开关或特定视角进行摄像的指令。

7、根据权利要求 4 所述的立体视频通信终端，其特征在于，所述发送装置进一步包括:

标定单元，用于获取所述摄像及图像处理单元中摄像机的内参和外参，并输出校正所述摄像机的指令到所述采集控制单元。

8、根据权利要求 4 所述的立体视频通信终端，其特征在于，所述发送装置进一步包括:

预处理单元，用于接收所述采集控制单元输出的视频数据和相应的摄像机参数，并根据预处理算法对所述视频数据进行预处理。

9、根据权利要求 4 所述的立体视频通信终端，其特征在于，所述发送装置进一步包括:

同步单元，用于产生同步信号并输入到所述摄像及图像处理单元，控制其进行同步采集; 或将所述同步信号输入到所述采集控制单元，通知其控制所述多摄像及图像处理单元进行同步采集。

10、根据权利要求 1 所述的立体视频通信终端，其特征在于，

所述发送装置进一步包括复用单元，用于对所述编码单元输出的编码数据进行复用，并输出到所述发送单元;

所述接收装置进一步包括解复用单元，用于对所述接收单元输出的复用数据进行解复用，并输出到所述解码单元。

11、根据权利要求 1 所述的立体视频通信终端，其特征在于，所述摄像及

图像处理单元是:

立体摄像及图像处理单元，输出包括深度和/或视差信息的视频数据；或

分开的摄像机和匹配/深度提取单元，所述摄像机用于摄像并输出视频数据，所述匹配/深度提取单元用于从所述摄像机输出的视频数据中得到成像物体的深度和/或视差信息并输出。

12、一种立体视频通信系统，其特征在于，包括:

立体视频通信终端，用于进行二维或立体视频通信;

二维视频通信终端，用于进行二维视频通信;

分组网络，用于承载所述立体视频通信终端之间、或二维视频通信终端之间传输的二维或立体视频数据。

13、根据权利要求 12 所述的立体视频通信系统，其特征在于，进一步包括:

多点控制系统，用于控制所述二维视频通信终端和/或立体视频通信终端间的多点会议连接，包括:

能力判断单元，用于在接收到所述通信终端的会议发起指令时，根据所述指令携带的能力信息判断会议双方是否具有立体摄像和立体显示能力;

会议建立单元，用于在所述能力判断单元判断双方都具有立体摄像和立体显示能力时，触发所述多点控制系统通过所述分组网络建立所述会议双方通信终端的会议连接。

14、根据权利要求 13 所述的立体视频通信系统，其特征在于，所述多点控制系统进一步包括:

转换单元，用于转换数据格式，包括将从一方终端接收到的视频数据转换成二维视频数据;

转发单元，用于将所述转换单元输出的二维视频数据发送给对端;

其中，所述多点控制系统的能力判断单元在进行能力判断的结果为会议双方有一个不具有立体显示能力时，启动所述转换单元工作。

15、根据权利要求 12 所述的立体视频通信系统，其特征在于，所述分组网络是有线分组网络或无线分组网络。

16、根据权利要求 15 所述的立体视频通信系统，其特征在于，所述有线

分组网络是局域网、E1、窄带综合业务数字网 ISDN 或宽带 ISDN。

17、根据权利要求 12 所述的立体视频通信系统，其特征在于，所述分组网络中进一步包括:

网守，提供所述分组网络各单元的地址转换和网络接入控制;

网关，为所述分组网络的通信双方、或者网关之间提供实时双向通信。

18、一种立体视频通信终端，其特征在于，包括:

摄像及图像处理单元，进行摄像，包括进行摄像并输出视频数据及其深度和/或视差信息;

编码单元，用于进行编码，包括对所述摄像及图像处理单元输出的视频数据及其深度和/或视差信息进行编码;

发送单元，用于将所述编码单元输出的编码数据封装成符合实时传输协议的数据包，并通过分组网络进行实时传输。

19、一种立体视频通信终端，其特征在于，包括:

接收单元，用于接收数据包，包括接收发送单元发送的数据包并去除所述数据包的协议头，得到所述编码数据;

解码单元，用于解码所述接收单元输出的编码数据，得到所述视频数据以及深度和/或视差信息;

重构单元，用于根据所述解码单元输出所述深度和/或视差信息、以及所述解码单元输出的视频数据进行用户所处视角图像的重构，并输入到所述渲染单元;

渲染单元，用于将所述重构单元输出的重构图像数据渲染到立体显示设备上。

20、根据权利要求 19 所述的立体视频通信终端，其特征在于，进一步包括:

转换单元，用于将解码单元输出的立体视频数据转换成二维视频数据;

平面显示设备，用于显示所述转换单元输出的二维视频数据。

21、一种立体视频通信方法，其特征在于，进行双向立体视频通信，包括:

进行摄像，得到视频数据;

从所视频数据中得到成像物体的深度和/或视差信息;

对所述视频数据和深度和/或视差信息进行编码;

将所述编码数据封装成符合实时传输协议的数据包;

通过分组网络发送所述数据包。

22、根据权利要求 21 所述的立体视频通信方法，其特征在于，所述进行摄像，得到视频数据包括:

进行多视角摄像，得到多视角视频数据。

23、根据权利要求 21 所述的立体视频通信方法，其特征在于，

所述进行双向立体视频通信还包括: 发送携带摄像及图像处理单元能力信息的会议发起指令;

通过分组网络发送所述数据包之后进一步包括:

根据接收到的所述会议发起指令及其携带的能力信息，判断会议双方是否具有立体摄像和立体显示能力;

在判断双方都具有立体摄像和立体显示能力时，通过所述分组网络建立所述会议双方通信终端的会议连接，启动通信双方的摄像及图像处理单元和接收装置。

24、根据权利要求 21 所述的立体视频通信方法，其特征在于，所述步骤: 进行摄像，得到视频数据包括:

获取摄像机的内参和外参，并根据所述内参和外参校正所述摄像操作。

25、一种立体视频通信方法，其特征在于，包括:

接收视频数据，包括:

接收通过分组网络进行实时传输的视频数据包，并去除所述数据包的协议头，得到已编码立体视频编码数据;

对所述已编码视频数据进行解码，得到视频数据以及相应的深度和/或视差信息;

根据所述深度和/或视差信息、以及视频数据进行用户所处视角图像的重构;

将所述重构的图像数据渲染到立体显示设备上。

26、根据权利要求 25 所述的立体视频通信方法，其特征在于，对所述已编码视频数据进行解码之后，进一步包括:

判断本端显示设备是否具有立体显示能力，在判断为否时，将解码后的立体视频数据转换成二维视频数据并向平面显示设备发送。

27、根据权利要求 25 所述的立体视频通信方法，其特征在于，在去除所述数据包的协议头之后，解码之前，进一步包括:

判断所述数据包是否是复用的视频数据，在判断为是时对所述数据包进行解复用。

28、根据权利要求 25 所述的立体视频通信方法，其特征在于，在渲染到立体显示设备上之前，进一步包括:

判断所述已解码数据是否需要进行图像重构;

在判断为需要进行重构时，对所述已解码数据进行图像重构。

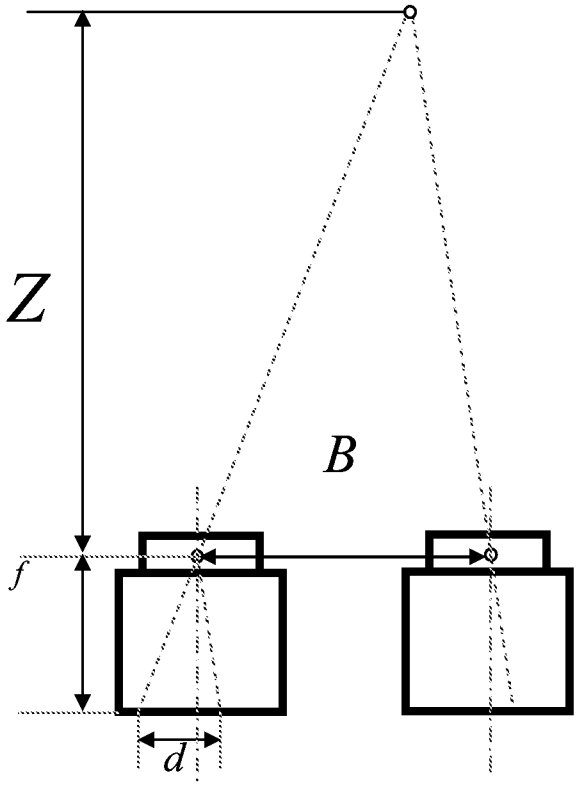


图 1

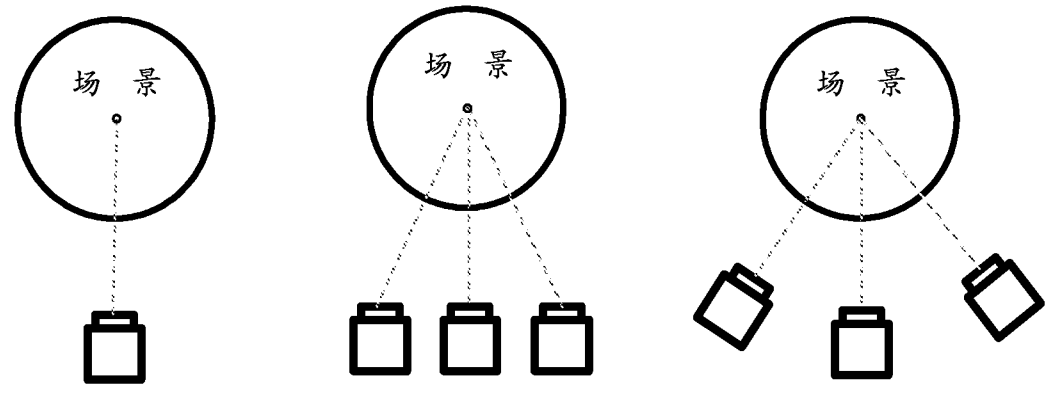


图 2

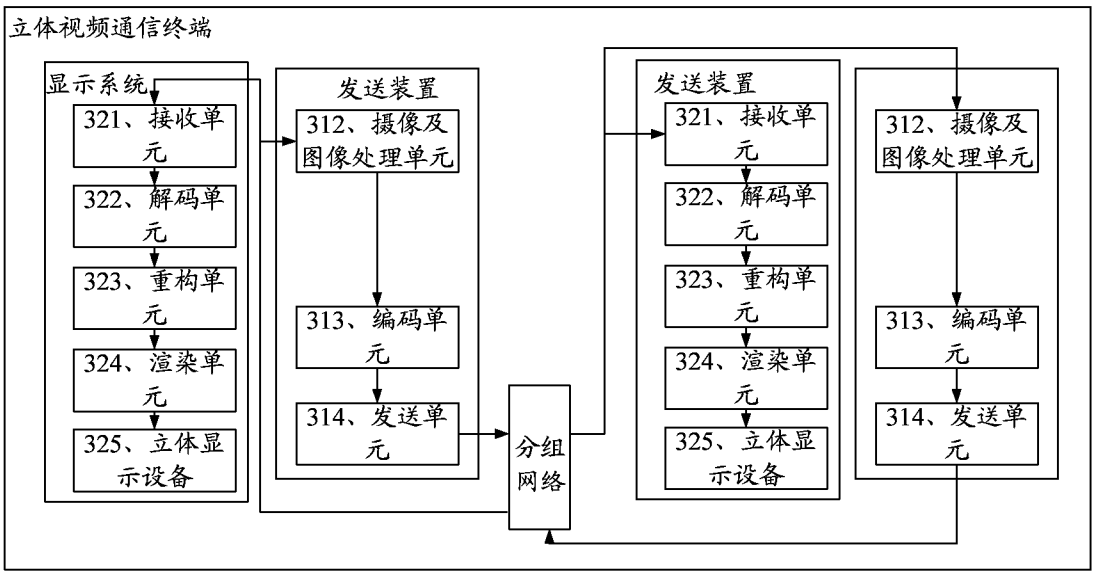

立体视频通信终端
显示系统
321、接收单元
322、解码单元
323、重构单元
324、渲染单元
325、立体显示设备
发送装置
312、摄像及图像处理单元
313、编码单元
314、发送单元
分组网络
发送装置
321、接收单元
322、解码单元
323、重构单元
324、渲染单元
325、立体显示设备
312、摄像及图像处理单元
313、编码单元
314、发送单元

图 3

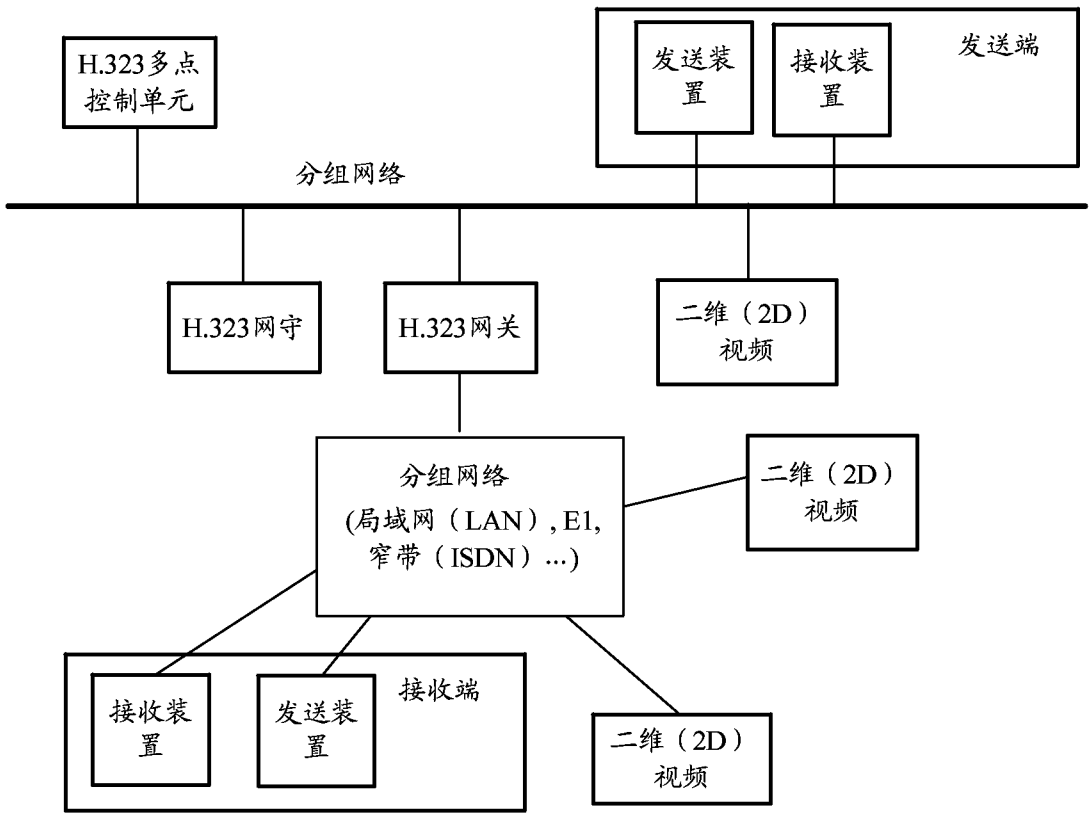

H.323多点控制单元
发送装置
接收装置
发送端
分组网络
H.323网守
H.323网关
二维（2D）视频
分组网络
(局域网（LAN）, E1, 窄带（ISDN）...)
二维（2D）视频
接收装置
发送装置
接收端
二维（2D）视频

图 4

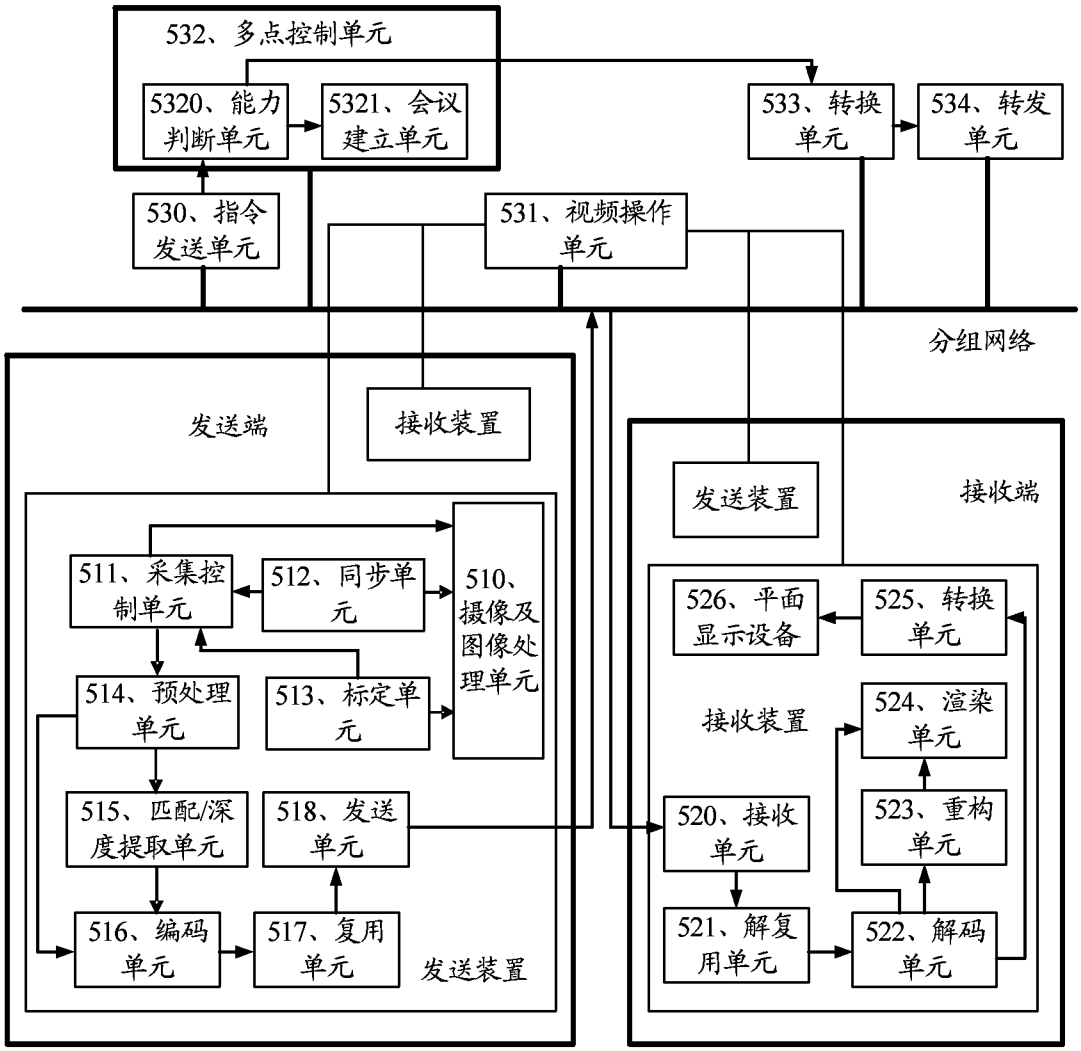

532、多点控制单元
5320、能力判断单元
5321、会议建立单元
533、转换单元
534、转发单元
530、指令发送单元
531、视频操作单元
分组网络
发送端
接收装置
511、采集控制单元
512、同步单元
510、摄像及图像处理单元
514、预处理单元
513、标定单元
515、匹配/深度提取单元
518、发送单元
516、编码单元
517、复用单元
发送装置
发送装置
接收端
526、平面显示设备
525、转换单元
接收装置
524、渲染单元
520、接收单元
523、重构单元
521、解复用单元
522、解码单元

图 5

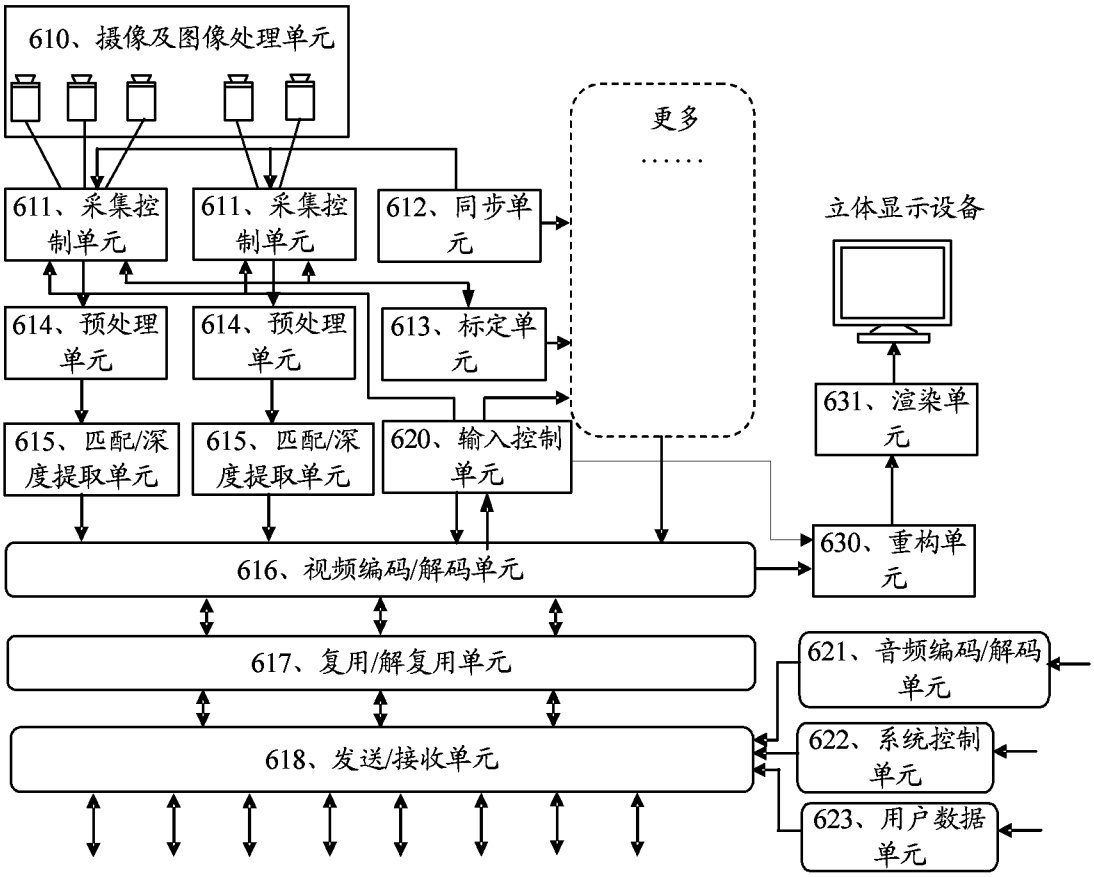

610、摄像及图像处理单元
更多
……
611、采集控制单元
611、采集控制单元
612、同步单元
立体显示设备
614、预处理单元
614、预处理单元
613、标定单元
631、渲染单元
615、匹配/深度提取单元
615、匹配/深度提取单元
620、输入控制单元
630、重构单元
616、视频编码/解码单元
617、复用/解复用单元
621、音频编码/解码单元
618、发送/接收单元
622、系统控制单元
623、用户数据单元

图 6

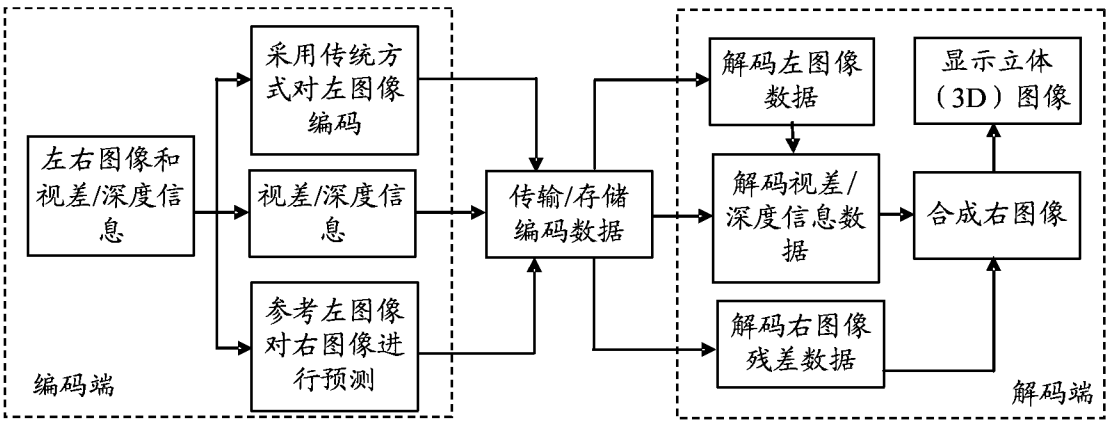

左右图像和视差/深度信息
采用传统方式对左图像编码
视差/深度信息
参考左图像对右图像进行预测
编码端
传输/存储编码数据
解码左图像数据
解码视差/深度信息数据
解码右图像残差数据
显示立体（3D）图像
合成右图像
解码端

图 7

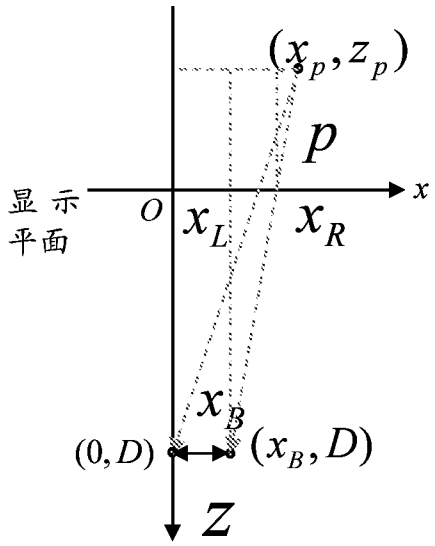

(x_p, z_p)
p
显示平面
O
x
x_L
x_R
x_B
$(0, D)$
(x_B, D)
Z

图 8

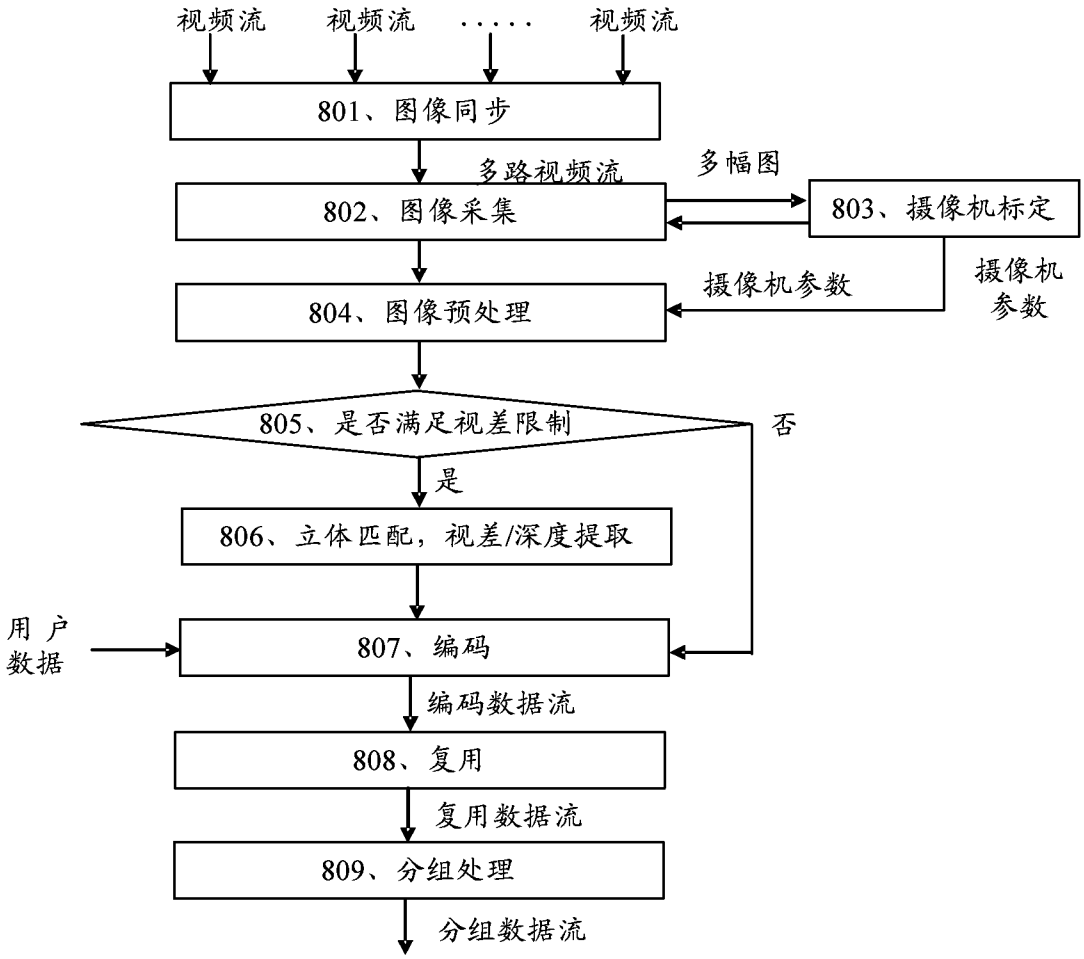

视频流
视频流
.....
视频流
801、图像同步
多路视频流
多幅图
802、图像采集
803、摄像机标定
摄像机参数
摄像机参数
804、图像预处理
805、是否满足视差限制
否
是
806、立体匹配，视差/深度提取
用户数据
807、编码
编码数据流
808、复用
复用数据流
809、分组处理
分组数据流

图 9

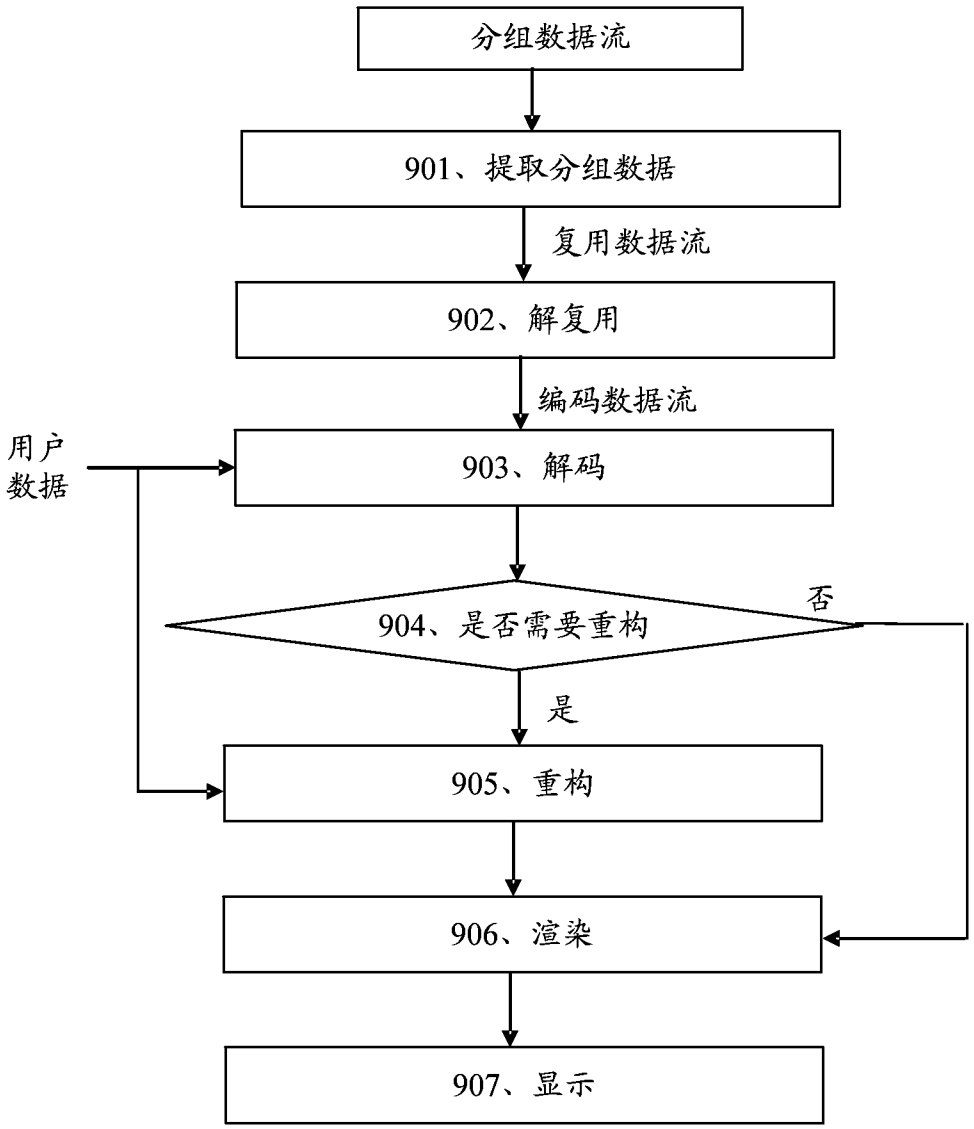

分组数据流
901、提取分组数据
复用数据流
902、解复用
编码数据流
用户数据
903、解码
904、是否需要重构
否
是
905、重构
906、渲染
907、显示

图 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2008/073310

A. CLASSIFICATION OF SUBJECT MATTER

H04N 13/00 (2006.01)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 13/00; H04N 7/-

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)

WPI; EPODOC; PAJ: three dimensional，two dimensional，3D，2D，solid，video，communication，send，receive，terminal，interactive，network，transmit

CNPAT; CNKI:

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP2005-25388A (TOPPAN PRINTING CO LTD) 27 January 2005（27.01.2005）description segment 0027- segment 0032，figure 1	12，15-17
Y	CN101013507A (NEC CORP) 08 August 2007（08.08.2007）description page14，line11-line19，figure 16	12，15-17
A	CN1275871A (UNIV QINGHUA) 06 December 2000（06.12.2000）the whole document	1-28

☐ 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
02 March 2009（02.03.2009）	**12 Mar. 2009 (12.03.2009)**
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer WU Shuang Telephone No. (86-10)62411507

Form PCT/ISA/210 (second sheet) (April 2007)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2008/073310

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP2005-25388A	27.01.2005	None	
CN101013507A	08.08.2007	US2007195082A1	23.08.2007
		JP2007200227A	09.08.2007
CN1275871A	06.12.2000	CN1134175C	07.01.2004

国际检索报告

国际申请号
PCT/CN2008/073310

A. 主题的分类

H04N 13/00 (2006.01)i

按照国际专利分类表(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 13/00; H04N 7/-

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）)

WPI; EPODOC; PAJ: three dimensional，two dimensional，3D，2D，solid，video，communication，send，receive，terminal，interactive，network，transmit

CNPAT; CNKI: 三维，二维，立体，视频，通信，发送，接收，终端，交互，网络，3D，2D

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	JP2005-25388A (凸版印刷株式会社) 27.1 月 2005（27.01.2005）说明书第 0027 段-0032 段，附图 1	12，15-17
Y	CN101013507A (日本电气株式会社) 08.8 月 2007（08.08.2007）说明书第 14 页第 11 行一第 19 行，附图 16	12，15-17
A	CN1275871A (清华大学) 06.12 月 2000（06.12.2000）全文	1-28

☐ 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

"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
02.3 月 2009（02.03.2009）	**12.3 月 2009 (12.03.2009)**
中华人民共和国国家知识产权局(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 吴爽 电话号码：(86-10) **62411507**

国际检索报告

关于同族专利的信息

国际申请号
PCT/CN2008/073310

检索报告中引用的专利文件	公布日期	同族专利	公布日期
JP2005-25388A	27.01.2005	无	
CN101013507A	08.08.2007	US2007195082A1	23.08.2007
		JP2007200227A	09.08.2007
CN1275871A	06.12.2000	CN1134175C	07.01.2004