



(12) 发明专利申请

(10) 申请公布号 CN 103281589 A

(43) 申请公布日 2013. 09. 04

(21) 申请号 201310085001. 6

H04N 21/2365(2011. 01)

(22) 申请日 2008. 06. 19

H04N 21/434(2011. 01)

(30) 优先权数据

H04N 21/435(2011. 01)

10-2007-0101964 2007. 10. 10 KR

H04N 21/81(2011. 01)

10-2007-0119740 2007. 11. 22 KR

H04N 21/854(2011. 01)

10-2007-0122945 2007. 11. 29 KR

H04N 13/00(2006. 01)

10-2007-0133764 2007. 12. 18 KR

10-2008-0016280 2008. 02. 22 KR

(62) 分案原申请数据

200880119862. 9 2008. 06. 19

(71) 申请人 韩国电子通信研究院

地址 韩国大田市

(72) 发明人 尹国镇 曹叔嬉 许南湜 金镇雄

李寿寅

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 张银英

(51) Int. Cl.

H04N 21/4402(2011. 01)

H04N 21/235(2011. 01)

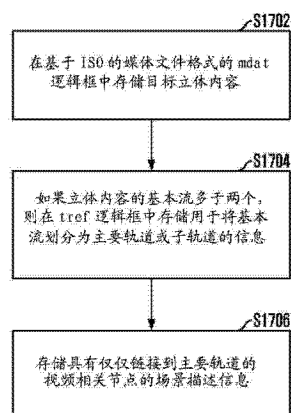
权利要求书2页 说明书32页 附图16页

(54) 发明名称

非瞬态性计算机可读存储介质

(57) 摘要

提供了一种用于存储和再现立体数据的元数据结构以及一种用于基于该元数据结构来存储立体内容文件的方法。所述立体内容文件存储方法包括:存储立体内容的流;以及,当立体内容具有多于两个基础流时,存储用于将基础流划分为主要轨道和补充轨道的信息。



1. 一种非瞬态性计算机可读存储介质,包括用于执行如下操作的编程指令:
接收用于形成立体内容的第一基础流和第二基础流的参考信息;以及
在包括用于存储所述参考信息的参考信息区域的文件格式中存储参考信息,
其中,所述参考信息包括用于定义第一基础流和第二基础流的参考关系的关系信息,
所述关系信息包括被标识为主要轨道和补充轨道的轨道。
2. 根据权利要求1的非瞬态性计算机可读存储介质,其中,所述文件结构还包括用于存储所述第一基础流和所述第二基础流的内容区域。
3. 根据权利要求1的非瞬态性计算机可读存储介质,其中,所述文件结构是基于了基于国际标准化组织(ISO)的媒体文件格式的结构。
4. 根据权利要求1的非瞬态性计算机可读存储介质,其中,所述第一基础流和第二基础流包括单视场片段,并且其中所述参考信息包括用于定义当显示单视场片段时、在第一基础流和第二基础流之间的要显示的基础流的信息。
5. 根据权利要求1的非瞬态性计算机可读存储介质,其中,将所述立体内容形成为左侧视图序列和右侧视图序列,并且其中所述左侧视图序列形成第一基础流,而所述右侧视图序列形成第二基础流。
6. 根据权利要求1的非瞬态性计算机可读存储介质,还包括用于执行如下操作的编程指令:
接收包括立体内容的片段信息的立体视频媒体信息;以及
在还包括用于存储立体视频媒体信息的视频媒体信息区域的文件结构中存储立体视频媒体信息。
7. 根据权利要求1的非瞬态性计算机可读存储介质,还包括用于执行如下操作的编程指令:
接收用于立体内容的立体相机和显示信息;以及
在还包括用于存储立体相机和显示信息的相机和显示信息区域的文件结构中存储立体相机和显示信息。
8. 根据权利要求7的非瞬态性计算机可读存储介质,其中,所述立体相机和显示信息包括:用于提供关于立体片段的相机和显示信息的立体片段的数目、用于标识立体片段的项目ID、关于是否存在项目ID的参考的信息之中的至少一个。
9. 根据权利要求1的非瞬态性计算机可读存储介质,还包括用于执行如下操作的编程指令:
存储包括与所述第一基础流和第二基础流之间的预定流链接的视频相关节点的、用于立体内容的场景描述信息。
10. 根据权利要求9的非瞬态性计算机可读存储介质,其中,所述视频相关节点链接到与预定流对应的ID或媒体流。
11. 一种非瞬态性计算机可读存储介质,具有:
文件格式结构,用于存储立体内容;
媒体数据逻辑框,用于存储立体内容的流;以及
轨道参考逻辑框,用于当立体内容包括多于两个基础流时,存储用于将基础流划分为主要轨道和补充轨道的信息,以及

用于执行如下操作的编程指令：通过从媒体数据逻辑框的补充轨道中去除冗余立体内容、并且在轨道参考逻辑框中包括对主要轨道中的冗余立体内容的参考，来去除在所述主要轨道和所述补充轨道中包括的立体内容的冗余。

12. 一种用于基于高级视频编码(AVC)补足增强信息(SEI)来存储立体内容的非瞬态性计算机可读存储介质，所述非瞬态性计算机可读存储介质包括用于执行如下操作的指令：

存储基于 AVC 而编码的立体内容，

其中“立体视频信息 SEI”包括指示立体视频是“并排”类型的信息，以及

基于“reserved_sei_message”来存储对于立体内容的每个流所需的相机参数和显示信息，其中，通过从第一轨道中去除冗余显示信息并且包括对包含于第二轨道中的所述冗余显示信息的参考，来去除在所述立体内容中包括的立体内容的冗余。

13. 根据权利要求 12 的非瞬态性计算机可读存储介质，其中所述相机参数和显示信息包括：是否存在显示参数的信息、是否存在相机参数的信息、关于相机运动的信息、关于从镜头到图像平面的距离的信息、关于从基线到会聚点的距离的信息、和关于相机安排的信息之中的至少一个，并且包括关于左侧和右侧图像之间最大视差和最小视差的信息。

14. 根据权利要求 1 的非瞬态性计算机可读存储介质，还包括用于执行如下操作的指令：在基于国际标准化组织(ISO)的文件格式中定义使得第二基础流能够参考第一显示轨道中的所述冗余显示信息的逻辑框。

非瞬态性计算机可读存储介质

[0001] 本申请是申请日为 2008 年 6 月 19 日、申请号为 200880119862.9、发明名称为“用于存储和播放立体数据的元数据结构、以及用于使用该元数据来存储立体内容文件的方法”的发明专利申请的分案申请。

技术领域

[0002] 本发明涉及一种用于存储和再现立体数据的元数据结构、以及用于使用该元数据结构来存储立体内容文件的方法；且更具体地，涉及一种当一起提供二维(2D)和三维(3D)内容时所需的元数据结构(控制信息)、以及用于使用所述元数据结构来存储立体内容文件的方法。

[0003] 该工作得到了 MIC/IITA 的 IT R&D 计划 [2005-S-403-02, “Development of Super-intelligent Multimedia Anytime-anywhere Realistic TV(SmarTV)Technology”] 的支持。

背景技术

[0004] 已经基于移动电话、数字相机、数字视频盘(DVD)、和 PDP 而在国内和国际上形成了使用立体内容的应用服务以及相关装置的市场。相应地，已经存在一种对于定义用于捕捉、存储、和再现立体内容的系统信息或控制信息(元数据)的标准以及包括所述标准的文件格式的需求。

[0005] 题目为“Apparatus and method for processing 3D moving images using MPEG-4 object descriptor and structure(用于使用 MPEG-4 对象描述符和结构来处理 3D 运动图像的设备和方法)”的韩国专利公布第 2006-0056070 号(在下文中，称为第一专利)公开了一种具有诸如 3D 运动图像的类型、各个显示类型、和视点之类的新信息的 3D 运动图像对象描述符。题目为 method for transmitting stereoscopic image data(用于传送立体图像数据的方法)的韩国专利申请第 2006-0100258 号(在下文中：第二专利)公开了一种包括视频数据单元和报头单元的文件格式、作为用于为了立体内容而解码的视频数据的文件格式，所述视频数据单元包括立体图像信息，并且所述报头单元包括用于解码和再现立体图像信息的元数据。

[0006] 然而，所述第一和第二专利未能引入一种用于当一起组织和提供 2D 内容或 3D 内容时(即，当一起使用 2D 内容和 3D 内容时)标识内容的方法、一种用于提供当 3D 内容具有不同的立体相机和显示信息时的立体相机和显示信息、以及当 3D 内容由两个基础(elementary)流形成时的立体轨道参考信息的方法。

发明内容

[0007] 技术问题

[0008] 本发明的实施例旨在提供一种用于存储立体内容的方法。

[0009] 本发明的另一实施例旨在提供一种用于当在各个 3D 终端环境中下载和再现 2D 内

容和 3D 内容时通过标识 2D 内容和 3D 内容来存储立体内容、并且通过自动地接通 / 关断视差屏障而显示所标识的 2D 内容和 3D 内容的方法。

[0010] 本发明的又一实施例旨在提供一种用于当立体内容具有不同的立体相机和显示信息时、通过片段来存储立体相机 / 显示信息的方法。

[0011] 本发明的又一实施例旨在提供一种用于当立体内容具有两个基础流时存储用于标识主要 / 子轨道的立体内容、并且在维持与典型 2D 终端处的兼容性的同时去除在每个轨道中包括并使用的立体相机 / 显示信息的冗余的方法。

[0012] 本发明的其他目的和优点可以通过以下描述来理解, 并且通过参考本发明的实施例而变得明显。同样, 对于本发明的领域的技术人员明显的是, 可以通过所要求保护的部件及其组合来实现本发明的目的和优点。

[0013] 技术方案

[0014] 根据本发明的一方面, 提供了一种用于存储立体内容的方法, 所述方法包括: 存储立体内容的流; 以及当立体内容具有多于两个基础流时, 存储用于将基础流划分为主要轨道和补充轨道的信息。

[0015] 根据本发明的另一方面, 提供了一种用于存储立体内容的文件格式结构, 所述文件格式结构包括: 媒体数据逻辑框(box), 用于存储立体内容的流; 以及轨道参考逻辑框, 用于当立体内容包括多于两个基础流时, 存储用于将基础流划分为主要轨道和补充轨道的信息。

[0016] 根据本发明的另一方面, 提供了一种用于基于高级视频编码(AVC) 补足增强信息(SEI) 来存储立体内容的方法, 所述方法包括: 存储基于 AVC 而编码的立体内容, 其中“立体视频信息 SEI” 包括指示了立体视频是“并排”类型的信息。

[0017] 根据本发明的另一方面, 提供了一种非瞬态性计算机可读存储介质, 包括用于执行如下操作的编程指令: 接收用于形成立体内容的第一基础流和第二基础流的参考信息; 以及在包括用于存储所述参考信息的参考信息区域的文件格式中存储参考信息, 其中, 所述参考信息包括用于定义第一基础流和第二基础流的参考关系的关系信息, 所述关系信息包括被标识为主要轨道和补充轨道的轨道。

[0018] 根据本发明的另一方面, 提供了一种非瞬态性计算机可读存储介质, 具有: 文件格式结构, 用于存储立体内容; 媒体数据逻辑框, 用于存储立体内容的流; 以及轨道参考逻辑框, 用于当立体内容包括多于两个基础流时, 存储用于将基础流划分为主要轨道和补充轨道的信息, 以及用于执行如下操作的编程指令: 通过从媒体数据逻辑框的补充轨道中去除冗余立体内容、并且在轨道参考逻辑框中包括对主要轨道中的冗余立体内容的参考, 来去除在所述主要轨道和所述补充轨道中包括的立体内容的冗余。

[0019] 根据本发明的另一方面, 提供了一种用于基于高级视频编码(AVC) 补足增强信息(SEI) 来存储立体内容的非瞬态性计算机可读存储介质, 所述非瞬态性计算机可读存储介质包括用于执行如下操作的指令: 存储基于 AVC 而编码的立体内容, 其中“立体视频信息 SEI” 包括指示立体视频是“并排”类型的信息, 以及基于“reserved_sei_message” 来存储对于立体内容的每个流所需的相机参数和显示信息, 其中, 通过从第一轨道中去除冗余显示信息并且包括对包含于第二轨道中的所述冗余显示信息的参考来去除在所述立体内容中包括的立体内容的冗余。

[0020] 有益效果

[0021] 根据本发明,可以方便地存储和提供由 2D 内容和 3D 内容形成的立体内容。通过经由 2D 内容和 3D 内容定界符信息而将视差屏障自动地改变为 2D 或 3D 显示,来使得用户能够方便地享受立体内容。

附图说明

[0022] 图 1 是图示了用于提供立体内容的各个内容配置类型的图。

[0023] 图 2 是图示了当存在 MPEG-4 系统信息时的、根据本发明实施例的用于立体视频应用格式的基本文件格式的图。

[0024] 图 3 是图示了当不存在 MPEG-4 系统信息时的、根据本发明实施例的用于立体视频应用格式的基本文件格式的图。

[0025] 图 4 是图示了当存在 MPEG-4 系统信息时的、根据本发明第一实施例的立体视频应用格式的图。

[0026] 图 5 是图示了当不存在 MPEG-4 系统信息时的、根据本发明第一实施例的立体视频应用格式的图。

[0027] 图 6 是图示了根据本发明实施例的“ishd”逻辑框被包括在“moov”逻辑框中的存储格式的图。在图 6 中,图 a) 示出了用于由一个源形成的 3D 内容的格式,而图 b) 示出了用于由两个源形成的 3D 内容的格式。

[0028] 图 7 是图示了根据本发明实施例的“ishd”逻辑框被包括在“mdat”逻辑框中的存储格式的图。在图 7 中,图 a) 示出了由一个源形成的 3D 内容,而图 b) 示出了由两个源形成的 3D 内容。

[0029] 图 8 是图示了根据本发明实施例的包括“ishd”和“meta”的存储格式的图。在图 8 中,图 a) 示出了由一个源形成的 3D 内容,而图 b) 示出了由两个源形成的 3D 内容。

[0030] 图 9 是图示了根据本发明实施例的包括“ishd”逻辑框和 LAsEr 的格式的图。在图 9 中,图 a) 示出了包括在“moov”逻辑框中包括的“ishd”逻辑框和 LAsEr 的格式,图 b) 示出了“ishd”逻辑框被包括在“moov”逻辑框中并且 LAsEr 被包括在“mdat”逻辑框中的格式。

[0031] 图 10 是根据本发明实施例的当 ES=1 时包括“ishd”逻辑框和“iloc”逻辑框的 SS-VAF 的图。

[0032] 图 11 是根据本发明实施例的包括“ishd”逻辑框和“iloc”逻辑框的 SS-VAF 的图。

[0033] 图 12a) 示出了在 AVC 中的 SEI 的基本结构中的包括“立体视频信息 SEI (stereo video information SEI)”和“reserved_sei_message (保留的 sei 消息)”的部分,而图 12b) 示出了 SEI 在 AVC 流中的位置。

[0034] 图 13 是图示了根据本发明实施例的使用“立体视频信息 SEI”和“reserved_sei_message”的 SS-VAF 的图。

[0035] 图 14 是图示了当 ES=2 时仅仅由单一格式立体流形成内容的情况下的、根据本发明实施例的具有“tref”、“ishd”、和“iloc”逻辑框的 SS-VAF 的图。

[0036] 图 15 是图示了根据本发明实施例的具有“tref”、“ishd”、和“iloc”逻辑框的

SS-VAF 的图。

[0037] 图 16 是图示了根据本发明实施例的用于存储立体内容的方法的流程图。

[0038] 图 17 是图示了根据本发明另一实施例的用于存储立体内容的方法的流程图。

[0039] 图 18 是图示了根据本发明又一实施例的用于存储立体内容的方法的流程图。

具体实施方式

[0040] 图 1 图示了用于提供立体内容的各个内容配置。图 1 的图 a) 示出了用于具有一个基础流 ES (ES=1) 的内容的格式, 而图 1 的图 b) 示出了用于具有两个基础流 ES (ES=2) 的内容的格式。单一格式是当立体图像形成方案相同时并且当仅仅包括一个相机参数和一个显示信息时的用于内容的格式。多个格式是当立体图像形成方案不同时、当尽管立体图像形成方案相同但是相机参数不同时、或当包括多个显示信息或另一信息时的格式。

[0041] Stereoscopic_Content_Type 包括: i) 诸如双目 3D 视频服务之类的立体视频内容、ii) 诸如双目 3D 静止图像服务(例如: 幻灯片)、2D (单) 视频和 3D 数据服务(预定的场景或部分) 的组合之类的立体图像内容、iii) 诸如 2D (单) 视频和 3D 视频(预定的场景或部分) 的组合之类的单视场(monoscopic) 和立体混合内容。

[0042] 图 2 和 3 示出了根据本发明实施例的立体视频应用格式(SS-VAF) 的基本结构。

[0043] 图 2 是包括 MPEG-4 系统信息的文件格式结构, 而图 3 是不包括 MPEG-4 系统信息的文件格式结构。如图所示的, SS-VAF 包括“ftyp”逻辑框、“moov”逻辑框、和“mdat”逻辑框。在下文中, 将描述 SS-VAF 的逻辑框的语法和语义。在 SS-VAF 的结构中包括根据本实施例的逻辑框, 可以根据其类型来改变其位置, 并且可以独立地使用在所述逻辑框中包括的信息。

[0044] 1. “scty” (立体内容类型)

[0045] “scty”表示内容的基本类型。即, 通过“scty”来将内容分类为单视场内容或立体内容。这里, 单视场内容意味着一般的 2D 图像。表 1 示出了“scty”的语法。可以在“ftyp”逻辑框或另一逻辑框中包括“scty”中的“Stereoscopic_Content_Type”。

[0046] 表 1

[0047]

```
Aligned(8) class StereoscopicContentTypeBox extend Box('sfty')
{
  unsigned int(2) Stereoscopic_Content_Type:
    unsigned int(6) reserved;
}
```

[0048] 在表 1, “Stereoscopic_Content_Type”表示立体内容类型, 并且具有表 2 的含义。

[0049] 表 2

[0050]

值	描述
00	立体内容 (3D)
01	单视场 / 立体混合内容 (2D+3D)
10 ~ 11	保留

[0051] 2. “sovf” (立体对象视觉格式)

[0052] “sovf”表示立体内容的图像配置格式(或视觉格式)。表 3 示出了“sovf”的语法。“sovf”包括“Stereoscopic_Object_VisualFormat”。可以在典型的其他逻辑框或为了存储立体内容而新定义的逻辑框中包括“Stereoscopic_Object_VisualFormat”。

[0053] Table3

[0054]

```
Aligned(8) class StereoscopicObjectDataVisualFormat extend Box('sovf')
{
    unsigned int(4) Stereoscopic_Object_VisualFormat;
    unsigned int(4) reserved;
}
```

[0055]

[0056] 在表 3 中,“Stereoscopic_Object_VisualFormat”表示立体内容的图像配置信息,并且具有表 4 的含义。

[0057] 表 4

[0058]

值	描述
0000	并排(side by side)
0001	从上到下
0010	基于水平方向的场序
0011	基于垂直方向的场序
0100	帧序
0101 ~ 0111	保留
1000	主要 + 附加(完全尺寸)
1001	主要 + 附加(垂直方向的一半)
1010	主要 + 附加(水平方向的一半)
1011	主要 + 附加(垂直方向 / 水平方向的一半)
1100	主要 + 深度图

[0059] 在表 4 中,“完全尺寸”意指补充图像的尺寸与主要图像的尺寸相同。“垂直方向的一半”表示补充图像的尺寸在垂直方向上是主要图像的尺寸的一半。“水平方向的一半”表示补充图像的尺寸在水平方向上是主要图像的尺寸的一半。“垂直方向 / 水平方向的一半”意指补充图像的尺寸在水平方向和垂直方向上是主要图像尺寸的一半。

[0060] 3. “ssci”(立体内容信息)

[0061] “ssci”表示关于立体内容的最小 / 最大深度或视差(disparity)的信息。“ssci”使得 3D 终端能够再现适合于 3D 显示的 3D 图像。表 5 表示了“ssci”的语法。可以在其他典型的逻辑框或为了存储立体内容而新定义的逻辑框中包括在“ssci”中包括的最小 / 最大深度信息。在表 5 中,“Max_of_depth(视差)”表示最大深度 / 视差信息(像素单位),而“Min_of_depth(视差)”表示最小深度 / 视差信息(像素单位)。

[0062] 表 5

[0063]


```

Aligned(8) class StereoscopicObjectContentInformation extend Box('ssci') {
    unsigned int(32) Max_of_depth or disparity;
    unsigned int(32) Min_of_depth or disparity;
}

```

[0064] 4. “scpi” (立体相机参数信息)

[0065] “scpi”表示由立体相机捕捉的或者由相关的工具创建的立体内容的相机参数信息。表 6 示出了“scpi”的语法。同样,可以在其他典型的逻辑框或为了存储立体内容而新定义的逻辑框中包括在“scpi”中包括的每个字段。

[0066] 表 6

[0067]

```

Aligned(8) class StereoscopicCameraParameterInformation extend Box('scpi') {
    unsigned int(32) Baseline;
    unsigned int(32) Focal_Length;
    unsigned int(32) Convergence_point_distnace;
    unsigned int(1) StereoscopicCamera_setting;
    unsigned int(7) Reserved;
}

```

[0068] 在表 6 中,“基线(Baseline)”表示左侧相机和右侧相机之间的距离,“Focal_Length”意指图像平面(CCD 传感器)和相机中心之间的距离,而“Convergence_point_distance”表示会聚点和基线之间的距离。这里,会聚点意指左侧相机和右侧相机的视觉线的交叉点。“StereoscopicCamera_setting”表示立体摄影 / 数据的相机安排,并且具有表 7 的含义。

[0069] 表 7

[0070]

值	描述
0	平行安排
1	交叉安排

[0071] 5. “iods” (对象描述符逻辑框)

[0072] “iods”表示关于初始对象描述符(IOD)的信息,所述初始对象描述符用于当包括诸如 BIFS 或 LAsEr 之类的场景信息时、表现 BIFS 流或 OD 流的位置。表 8 示出了“iods”的语法。

[0073] 表 8

[0074]

```

Aligned(8) class ObjectDescriptoratBox extend Box('iods') {
    ObjectDescriptor OD;
}

```

[0075] 6. “soet” (立体一个 ES 类型逻辑框)

[0076] “soet”表示当编码器输出一个基础流(ES)时的ES的类型。表9示出了“soet”的语法。

[0077] 表9

[0078]

```
Aligned(8) class StereoscopicOnesTypeBox extend Box('soet') {
    unsigned int(3) Stereoscopic_OneES_Type;
    unsigned int(5) Rreserved;
}
```

[0079] 在表9中,“Stereoscopic_OneES_Type”表示由一个ES形成的立体数据的图像配置格式的实施例,并且具有表10的含义。

[0080] 表10

[0081]

值	描述
000	并排
001	从上到下
010	基于水平方向的场序
011	基于垂直方向的场序
100	帧序
101	单视场 / 立体混合数据
110	基于补充数据而形成的参考图像 / 立体数据
111	保留

[0082]

[0083] 7. “stet” (立体两个ES类型逻辑框)

[0084] “stet”表示当编码器输出两个基础流(ES)时的每个ES的类型。表11表示了“stet”的语法。

[0085] 表11

[0086]

```
Aligned(8) class StereoscopicTwosTypeBox extend Box('stet') {
    unsigned int(2) Stereoscopic_TwoES_Type;
}
```

[0087] 在表11中,“Stereoscopic_TwoES_Type”表示由两个ES形成的立体数据的图像配置格式的实施例,并且具有表12的含义。

[0088] 表12

[0089]

值	描述
000	参考图像
001	补充数据（具有参考图像的相同尺寸的图像）
010	补充数据（具有参考图像的垂直方向上一半尺寸的图像）
011	补充数据（具有参考图像的水平方向上一半尺寸的图像）
100	补充数据（具有参考图像的水平方向和垂直方向上一半尺寸的图像）
101	补充数据（深度图）
110	补充数据（视差图）
111	保留

[0090] 8. “sstt”（立体时间表逻辑框）

[0091] “sstt”表示关于当一起使用单视场内容和立体内容时、场景中的单视场内容和立体内容的开始和结束的信息。表 13 示出了根据本发明第一实施例的“sstt”的语法。可以在其他典型的逻辑框或为了存储立体内容而新定义的逻辑框中包括“sstt”的每个字段。

[0092] 表 13

[0093]

```

Aligned(8) class StereoscopicTimeTable Box extend Box('sstt') {
    int i;
    unsigned int(8) Mono/stereoscopic_Scene_count;
    for(i=0;i<=Mono/stereoscopic_Scene_count;i++)
    {
        unsigned int(4) Mono/stereoscopic_identifier;
        unsigned int(4) Reserved;
        unsigned int(32) Start_Time;
    }
}

```

[0094] 在表 13 中，“Mono/stereoscopic_Scene_count”表示当由 2D 内容和 3D 内容一起形成立体内容时单视场 / 立体场景改变的数目。即，如果立体内容由 2D → 3D → 2D 形成，则将“Mono/stereoscopic_Scene_count”设置为 2。同样，如果立体内容仅仅由 3D 内容形成，而没有 2D 内容，则将“Mono/stereoscopic_Scene_count”设置为 2。这种信息可以用于 3D 终端中的 2D/3D 显示自动改变。

[0095] “Mono/stereoscopic_identifier”通过时间来表示内容类型，并且具有表 14 的含义。同样，“Mono/stereoscopic_identifier”可以用于标识 2D 内容或 3D 内容。例如，向“Mono/stereoscopic_identifier”分配 1 比特。如果“Mono/stereoscopic_identifier”是“0”，则“Mono/stereoscopic_identifier”表示 2D 内容。如果“Mono/stereoscopic_identifier”是“1”，则“Mono/stereoscopic_identifier”表示 3D 内容。“Start_Time”通

过时间来表示内容开始时间。

[0096] 表 14

[0097]

[0098]

值	描述
0000	单视场
0001	并排
0010	从上到下
0011	基于水平方向的场序
0100	基于垂直方向的场序
0101	帧序
0110 ~ 0111	保留
1000	主要 + 附加(完全尺寸)
1001	主要 + 附加(垂直方向的一半)
1010	主要 + 附加(水平方向的一半)
1011	主要 + 附加(垂直方向 / 水平方向的一半)
1100	主要 + 深度图
1101	主要 + 视差图

[0099] 表 15 示出了根据本发明第二实施例的“sstt”的语法。在表 15 中,“Start_Time”表示立体内容的开始时间,而“End_Time”表示立体内容的结束时间。

[0100] 表 15

[0101]

```
Aligned(8) class StereoscopicTimeTable Box extend Box('sstt') {
    int i;
    unsigned int(8) Stereoscopic_Scene_count;
    for(i=0;i<=Stereoscopic_Scene_count;i++)
    {
        unsigned int(32) Start_Time;
        unsigned int(32) End_Time;
    }
}
```

[0102] 表 16 示出了根据本发明第三实施例的“sstt”的语法。“Start_Sample_number”表示单视场 / 立体内容开始采样数目或采样的数目。即,采样的数目表示与单视场或立体对应的全部采样的数目。这里,采样意味着单独的视频帧和一系列时间邻接的视频帧。

[0103] 表 16

[0104]

```
Aligned(8) class StereoscopicTimeTable Box extend Box('sstt') {
    int i;
    unsigned int(16) Mono/stereoscopic_Scene_count;
    for(i=0;i<=Mono/stereoscopic_Scene_count;i++)
    {
        unsigned int(32) Start_Time;
        unsigned int(32) End_Time;
    }
}
```

[0105]

```
{
    unsigned int(4) Mono/stereoscopic_identifier;
    unsigned int(32) Start_Sample_number;
}
```

[0106] 表 17 示出了根据本发明第三实施例的“sstt”的语法。

[0107] 表 17

[0108]

```
Aligned(8) class StereoscopicTimeTable Box extend Box('sstt') {
    int i;
    unsigned int(8) Stereoscopic_Scene_count;
    for(i=0;i<=Stereoscopic_Scene_count;i++)
    {
        unsinged int(4) Stereoscopic_compositiontype;
        unsigned int(32) Start_sample_number;
        unsigned int(32) End_sample_number;
    }
}
```

[0109] 在表 17 中,当“Stereoscopic_compositiontype”由各个立体配置类型形成时,它通过时间来表示内容类型,并且具有表 18 的含义。“End_Sample_number”表示立体内容结束采样数目或采样的数目。

[0110] 表 18

[0111]

值	描述
0000	并排
0001	从上到下
0010	基于水平方向的场序
0011	基于垂直方向的场序
0100	帧序
0101 ~ 0111	保留
1000	主要 + 附加(完全尺寸)
1001	主要 + 附加(垂直方向的一半)
1010	主要 + 附加(水平方向的一半)
1011	主要 + 附加(垂直方向 / 水平方向的一半)
1100	主要 + 深度图
1101	主要 + 视差图

[0112]

[0113] 9. “sesn” (立体 ES Num 逻辑框)

[0114] “sesn”表示从编码器输出的基础流的数目。表 19 示出了“sesn”的语法。在表 19 中,“stereoscopic_ESNum”表示从用于立体内容的编码输出的基础流的数目。

[0115] 表 19

[0116]

```
Aligned(8) class StereoscopicESNumBox extend Box('seen') {
    unsigned int(16) stereoscopic_ESNum;
}
```

[0117] 10. “tref” (轨道参考逻辑框)

[0118] “tref”是在提供用于使得一个轨道能够参考其他轨道的信息的基于 ISO 的文件格式中定义的逻辑框。在“trak” (轨道逻辑框) 中包括“tref”。表 20 示出了根据本发明实施例的“tref”的语法。这里,“track_ID”表示要参考的轨道的标识。“reference_type”具有表 21 的含义。

[0119] 表 20

[0120]

```
aligned(8) class TrackReferenceBox extends Box ('tref') {}
aligned(8) class TrackReferenceTypeBox (unsigned int(32) reference_type)
extends Box(reference_type) {
    unsigned int(32) track_IDs[];
}
```

[0121] 表 21

[0122]

hint (提示)	所参考的 (多个) 轨道包含用于该提示轨道的原始媒体
cdsc	该轨道描述所参考的轨道
svdp	该轨道具有对于作为其主视图轨道的所参考轨道的依赖性, 并且它可以包含用于所参考的轨道的立体相关 “元 (meta)” 信息。

[0123]

[0124] 具有两个基础流 (ES) 的立体视频包括两个轨道, 并且连接到与作为场景描述的 LASer (如, 传统的 LASer) 中的视频相关的两个节点。即, 根据传统技术来将具有两个 ES 的立体视频识别为两个对象。然而, 因为将立体视频最终转换为一个 3D 视频格式并在终端中进行再现, 所以将立体视频识别为一个对象。即, 尽管使用两个轨道来形成立体视频, 但是应该仅仅使用一个节点来连接立体视频, 这是因为将立体视频转换为一个 3D 视频格式以再现场景。如果立体视频包括两个 ES, 则需要具有关于两个轨道之间的关系的信息, 并且在“tref”中的“svdp”处定义立体轨道参考信息 (如表 22 所示), 并且进行使用。尽管没有使用 LASer, 但是需要使用诸如“svdp”之类的立体轨道参考信息, 以用于存储具有两个 ES 的立体内容。

[0125] 如果立体内容包括两个基础流, 则“tref”使得能够将两个轨道标识为主要轨道和附加轨道。同样, “tref”使得附加轨道能够参考主要轨道, 并且仅仅在所述轨道之一中存储立体视频相关的立体相机和显示信息。因此, 可以消除信息的冗余。此外, 尽管立体内容包括两个轨道, 但是可以将一个轨道连接到 LASer 的一个视频节点。

[0126] 在本发明中,引入了初始立体报头(ishd)的结构,以便使得立体内容能够支持各个立体图像配置格式和相机参数。根据本发明的实施例,可以独立地使用在初始立体报头(ishd)中包括的信息。

[0127] 在立体流具有各个立体格式和相机参数的情况下,通过“iloc”来标识每个立体流以及每个立体流的开始和长度,并且向每个片段分配 item_ID。因此,初始立体报头(ishd)必须基于该 item_ID 来提供关于每个立体格式或相机参数的信息。这里,当立体流一起包括立体片段和单视场片段作为一个序列时,项目意指一个立体片段。

[0128] 如果立体流包括三个立体片段、并且每个立体片段包括不同的 ishd 信息,则通过 item_ID 及其描述信息来标识所述立体片段。然而,如果三个立体片段具有相同的 ishd 信息,则第二和第三立体片段包括示出第二和第三立体片段包括第一立体片段的相同 ishd 信息的字段。这种结构可以有利地去除在初始立体报头(ishd)中包括的信息冗余。

[0129] 图 4 是图示了根据本发明实施例的具有 MPEG-4 系统信息的立体视频应用格式(SS-VAF)的图,而图 5 是图示了根据本发明实施例的不具有 MPEG-4 系统信息的立体视频应用格式(SS-VAF)的图。

[0130] 当一起使用单视场内容和立体内容时,需要定界符信息来确定单视场内容或立体内容什么时候开始或结束。可以根据其中包括的单视场 / 立体内容的开始 / 结束信息、采样的 2D/3D 标识、和采样的数目(AU)来标识单视场内容和立体内容。

[0131] 图 6 到图 8 示出了用于支持像图 1 一样的各个内容配置格式的存储格式的构思结构。其基本结构包括“ftyp”逻辑框、“moov”逻辑框、和“mdat”逻辑框。“ftyp”逻辑框定义文件类型。即,“ftyp”逻辑框通过包括表现它是立体内容文件还是单视场 / 立体内容文件的字段来表现 3D 内容文件。“moov”逻辑框包括用于再现媒体数据的所有系统(元)信息,而“mdat”逻辑框包括实际的媒体数据。需要具有基于所示格式的用于立体内容的新补充信息,并且根据补充信息的位置来改变存储格式的结构。

[0132] 图 6 示出了包括在“moov”逻辑框中包括的初始立体报头(ishd)的存储格式的结构,所述初始立体报头具有关于形成了 3D 内容的源的数目的信息和新的补充信息。图 6 的图 a)示出了用于具有一个源的 3D 内容的存储格式。如图 a)所示,一个帧包括左侧图像信息和右侧图像信息两者(例如,并排)。图 6 的图 b)示出了用于具有两个源的 3D 内容的存储格式。如图 b)所示,在对应的帧中单独地包括左侧图像信息和右侧图像信息中的每一个。根据所包括的媒体数据的数目来改变“moov”逻辑框中的轨道的数目。“moov”逻辑框的轨道包括用于再现包括在“mdat”逻辑框中的媒体数据的整个系统信息(元信息)。

[0133] 这种存储格式需要一种用于新补充信息并且支持该新补充信息的结构。在本实施例中,新定义初始立体报头(ishd),并且将其包括在“moov”逻辑框的轨道中。可以在“moov”逻辑框中或在存储格式中改变初始立体报头(ishd)的位置。

[0134] 图 7 示出了具有含有关于新定义的初始立体报头的信息的“mdat”逻辑框的存储格式结构。图 7 的图 a)示出了用于由一个源形成的 3D 内容的存储格式,而图 7 的图 b)示出了用于由两个源形成的 3D 内容的存储格式。如所示的,可以在维持“moov”逻辑框的典型结构的同时、通过包括在“mdat”逻辑框中包括“ishd”流的信息来实施该存储格式。

[0135] 图 8 示出了包括具有 ishd 信息的“meta”逻辑框的存储格式。图 8 的图 a)示出了用于由一个源形成的 3D 内容的存储格式,而图 8 的图 b)示出了用于由两个源形成的 3D

内容的存储格式。

[0136] 表 22 示出了用于通知在“mdat”逻辑框中包括 ishd 信息的结构。在“stsd”（采样描述）逻辑框中包括这种结构。

[0137] 表 22

[0138]

//ishd序列

Class ishdSampleEntry(codingname) extends SampleEntry ('ishd') { }

[0139] 图 9 示出了基于图 6 到图 8 所示的结构的、具有关于场景描述符的信息的存储格式。场景描述符用于各个多媒体和与用户的交互的场景配置。在本实施例中，将 LAsER 用作场景描述符。

[0140] 图 9 的图 a) 示出了在“moov”逻辑框中包括用于存储场景描述符流的附加逻辑框的存储格式。图 9 的图 b) 示出了包括具有场景描述符流的“mdat”逻辑框、具有用于通知场景描述符流被包括在“mdat”逻辑框中的附加轨道的“moov”逻辑框、和具有关于场景描述符流的信息的“stsd”逻辑框的存储格式。即，它涉及在轨道中搜索“stsd”逻辑框、分析该轨道表现哪个信息（场景描述符 / 视频 / 音频）、和基于分析结果而使用在“mdat”逻辑框中存储的信息来进行解码。此外，还可以存在其它形式的存储格式结构，例如包括具有关于所定义的场景描述符的信息的“meta”逻辑框的存储格式结构。

[0141] 表 23 到表 25 示出了支持图 1 的所有 3D 内容配置格式的 ishd 结构的实施例。

[0142] 表 23

[0143]

Class ishd {

```
    unsigned int(16) num_MonoStereo_scene;  
    if(num_MonoStereo_scene){  
        for(i=0;i<num_MonoStereo_scene;i++)
```

```
        unsigned int(16) start_sample_index;  
        unsigned int(3) Composition_type;  
        unsigned int(1) numofES;
```

[0144]


```
        unsigned int(1) LR_first;
        unsigned int(3) Reserved;
    }
} else

    unsigned int(3) Composition_type;
    unsigned int(1) numofES;
    unsigned int(1) LR_first;
    stereoscopicCameraInfo[0...1];
    stereoscopicContentsInfo[0...1];
}
```

[0145] 表 24

[0146]

```
Class ishd {

    unsigned int(16) num_MonoStereo_scene;
    if(num_MonoStereo_scene){
        for(i=0;i<num_MonoStereo_scene;i++)

            unsigned int(16) start_sample_index;
            unsigned int(3) Composition_type;
            unsigned int(1) numofES;
            unsigned int(1) LR_first;
            unsigned int(3) Reserved;
        }
    } else

        unsigned int(3) Composition_type;
        unsigned int(1) numofES;
        unsigned int(1) LR_first;
    }
}
```

[0147] 表 25

[0148]

```
Class ishd {

    unsigned int(16) num_MonoStereo_scene;
```

[0149]

```
if(num_MonoStereo_scene){
    for(i=0;i<num_MonoStereo_scene;i++)

        unsigned int(16) start_sample_index;
        unsigned int(3)  Composition_type;
        unsigned int(1)  numofES;
        unsigned int(1)  LR_first;
        stereoscopicCameraInfo[0...1];
        stereoscopicContentsInfo[0...1];
        unsigned int(3)  Reserved;
    }
} else

    unsigned int(3)  Composition_type;
    unsigned int(1)  numofES;
    unsigned int(1)  LR_first;
    stereoscopicCameraInfo[0...1];
    stereoscopicContentsInfo[0...1];
}
```

[0150] 在表 23 到表 25 中,“num_MonoStereo_scene”表示当立体内容由 2D 内容和 3D 内容一起形成时的场景的数目。“num_MonoStereo_scene”还表示当立体内容由各个 3D 内容形成时的场景的数目。例如,立体内容由 2D 内容、3D 内容、和 2D 内容 [(2D) (3D) (2D)] 形成,num_MonoStereo_scene 变为 3 (Num_MonoStereo_scene=3)。如果以并排方案(场序)来形成立体内容,则 num_MonoStereo_scene 变为 2 (Num_MonoStereo_scene=2)。此外,如果立体内容以单一格式由 3D 内容形成,则 num_MonoStereo_scene 变为 1 (Num_MonoStereo_scene=1)。

[0151] 可以将“Start_sample_index”用作每个内容的开始采样数目(即,一般的帧数)或根据每个内容类型而包括的采样的数目。“numofES”表示在“mdat”逻辑框中包括的视频流的数目。

[0152] “Composition_type”表示用于标识 2D 内容和 3D 内容的格式的信息。可以将“Start_sample_index”和“Composition_type”用作用于支持 2D/3D 显示模式的各个 3D 终端处的自动显示开/关的基本信息。“Composition_type”具有表 26 的含义。

[0153] 表 26

[0154]

值		描述
3D	000	并排
	001	从上到下
	010	场序
	011	帧序
	100	垂直方向行交织格式
	101	左侧/右侧图像
2D	110	单视场左侧图像
	111	单视场右侧图像

[0155] “LR_first”表示左侧图像和右侧图像之间具有较高优先级的一个。即,“LR_first”通知左侧图像和右侧图像之间首先编码的图像。

[0156] “stereoscopicCameraInfo (立体相机信息)”对象表示用于 3D 内容的相机参数信息。表 27 示出了“stereoscopicCameraInfo”对象的实施例。可以在其他典型的逻辑框或根据本发明实施例而新定义的逻辑框中包括根据本实施例的相机参数信息。在表 27 中,“StereoscopicCamera_setting”表示当再现或摄影 3D 内容时的相机安排。即,“StereoscopicCamera_setting”表示“平行”和“交叉”之一。“Baseline(基线)”表示立体相机之间的距离,而“Focal_Length”表示从镜头到图像平面的距离。同样,“ConvergencePoint_distance”表示从连接了左侧相机和左侧相机的基线到会聚点的图。

[0157] 表 27

[0158]

<pre>StereoscopicCameraInfo { unsigned int(1) StereoscopicCamera_setting; unsigned int(7) Reserved=1111; unsigned int(16) Baseline; unsigned int(16) Focal_Length; unsigned int(16) ConvergencePoint_distance; }</pre>
--

[0159] “StereoscopicContentsInfo (立体内容信息)”对象表示用于显示 3D 内容的最小信息。表 28 示出了“StereoscopicContentsInfo”对象的实施例。可以在其他典型的逻辑框或根据本实施例而新定义的逻辑框中包括在“StereoscopicContentsInfo”中包括的信息。“Max_disparity”表示 3D 内容的最大视差尺寸,而“Min_disparity”表示 3D 内容的最小视差尺寸。

[0160] 表 28

[0161]

```
StereoscopicContentsinfo {  
    unsigned int(16) Max_disparity;  
    unsigned int(16) Min_disparity;  
}
```

[0162] 可以将“StereoscopicCameraInfo”和“StereoscopicContentsinfo”中的信息表达为诸如 MPEG-7 元数据之类的附加描述,并且进行存储。

[0163] 图 10 是图示了当 ES=1 时的根据本发明实施例的 SS-VAF 的图。

[0164] “ftyp”逻辑框表示是否包括立体内容。当整个基础流是 3D 时,并且当基础流由 2D/3D 混合流形成时,它们被认为是立体内容。

[0165] 当立体内容由 2D/3D 流形成时,需要 2D/3D 流的开始信息和长度信息。针对开始信息和长度信息,使用作为基于 ISO 的文件格式(11496-12)的典型逻辑框的“iloc”逻辑框。在立体内容的情况下,“iloc”逻辑框提供所存储文件中的立体片段的位置。

[0166] 通过“ishd”逻辑框来获得与区别 2D 流和 3D 流相关的信息。尽管在 2D/3D 混合流的情况下包括多个 3D 流,但是如果多个 3D 流是相同的信息(即,如果它是单一格式),则通过参考一个“ishd”信息来获得与 3D 流相关的信息。

[0167] 在立体内容仅仅由单一格式的 3D 流形成的情况下,可以使用“ishd”逻辑框、而不使用“iloc”逻辑框来表达立体数据。同样,使用“iloc”逻辑框来检测每个格式的偏移/长度值,并且如果当 ES=1 时、3D 流由多个格式形成,则通过“ishd”逻辑框来获得格式信息。在此情况下,“ishd”逻辑框包括关于多个格式的信息。

[0168] 图 11 是图示了当 ES=2 时的根据本发明实施例的 SS-VAF 的图。在 ES=2 的情况下,在对应的“trak”逻辑框中包括左侧流信息和右侧流信息。由于如上所述将立体数据转换为预定格式并进行显示,则将左侧流信息和右侧流信息形成为两个轨道。然而,需要示出左侧流和右侧流之间的关系,以便使得 将两个轨道识别为一个对象。例如,如果左侧图像是主要图像而右侧图像是附加图像,则可能通过表达在具有右侧图像流信息的“trak”逻辑框和具有左侧图像流信息的“trak”逻辑框之间的关系来去除“ishd”逻辑框中的冗余。如果在右侧图像流信息中包括的“ishd”中的相机参数和显示信息与在左侧图像流信息中包括的“ishd”中相同,则可以使用在左侧图像流信息中的“ishd”中包括的信息,而无需附加描述。为了表达这种关系,在本实施例中引入了“tref”逻辑框和“svdp”逻辑框。

[0169] 当与每个格式对应的 3D 流是由存储为左侧和右侧流的 3D 流上的多个格式的 3D 流形成时,需要检测它的定界符、开始、和长度。通过作为基于 ISO 的文件格式(14496-12)的典型逻辑框的“iloc”逻辑框来获得所述开始和长度信息。同样,通过“ishd”逻辑框来获得与多个格式的 3D 流的定界符相关的信息。表 29 示出了根据本发明实施例的单一格式的“ishd”逻辑框的语法。

[0170] 表 29

[0171]

```

Aligned(8) class Initial Stereoscopic Header box extend FullBox('ishd',
version=0,0)
    unsigned int(8) Stereoscopic_Composition_Type;
    unsigned int(1) LR_first;
    unsigned int(1) Is_camParams;
    unsigned int(1) Is_disInfo;
    unsigned int(5) Reserved;

    //所有以下是可选字段
    if(Is_CamParams){
        unsigned int(32) Baseline;
        unsigned int(32) focallength;
        unsigned int(32) convergence_distance;
        unsigned int(1) Is_camera_cross
    }

    if(Is_camera_cross){
        unsigned int(32) rotation[];
        unsigned int(32) translation[];
        unsigned int(7) reserved;
    }

    if(Is_disInfo){
        Int(16) MinofDisparity;
        Int(16) MaxofDisparity;
    }

```

[0172]

[0173] “Is_camParams”表示是否存在相机参数,“Is_disInfo”表示是否存在立体内容显示信息,“Bseline”表示左侧相机和右侧相机之间的距离,“focallength”表示从镜头到图像平面(底片)的距离,而“convergence_distance”表示从基线的中心到会聚点的距离。基线连接左侧相机和右侧相机,而会聚点是左侧和右侧相机的视线的交叉点。在平行轴相机的情况下,“convergence_distance”具有无穷大的值。为了表达它,向所有比特分派 1。

[0174] 同样,当“Is_camera_cross”是“1”时,“Is_camera_cross”表示交叉轴相机,而当“Is_camera_cross”是“0”时,“Is_camera_cross”表示平行轴相机。“rotation (旋转)”表示到对象的相机方位角。“translation (平移)”表示立体相机是否运动(当所有比特是 0 时,没有立体相机运动)。“MinofDisparity”表示左侧和右侧图像的最小视差尺寸,而“MaxofDisparity”表示左侧和右侧图像的最大视差。

[0175] 表 30 示出了多个格式的“ishd”逻辑框的语法。“Item_count”表示多个格式情

况下的格式的信息描述的数目。“Item_ID”表示每个格式的整体(integer)名称,并且用于标识立体内容中的各个立体格式。这里,与“iloc”逻辑框的 item_ID 一起使用该“Item_ID”。

[0176] 表 30

[0177]

```
Aligned(8) class Initial Stereoscopic Header box extend FullBox('ishd',  
version=0,0)  
  
    unsigned int(16)  item_count;  
    for (i=0; i<item_count; i++) {  
        unsigned int(16) item_ID;  
        unsigned int(8)  Stereoscopic_Composition_Type;
```

[0178]

```

unsigned int(1) LR_first;
unsigned int(1) Is_camParams;
unsigned int(1) Is_disInfo;
unsigned int(5) Reserved;

//所有以下是可选字段
if(Is_CamParams){
    unsigned int(32) baseline;
    unsigned int(32) focallength;
    unsigned int(32) convergence_distance;
    unsigned int(1) Is_camera_cross
}

if(Is_camera_cross) {
    unsigned int(32) rotation[];
    unsigned int(32) translation[];
    unsigned int(7) reserved;
}

if(Is_disInfo){
    Int(16) MinofDisparity;
    Int(16) MaxofDisparity;
}

//其他附加信息
if(other_flag){}
}

```

[0179] 在下文中,将描述高级视频编码(AVC)和补足增强信息(SEI)。SEI 包括具有与解码和显示相关的消息信息的“立体视频信息 SEI”,并且在 AVC 流内传送 SEI 消息。

[0180] 图 12 是包含 NAL 单元的单一视频基础流的流程图。图 12 的图 a)示出了包括“立体视频信息 SEI”和“reserved_sei_message”的部分,而图 12 的图 b) 示出了 AVC 流中的 SEI 的位置。表 31 示出了“立体视频信息 SEI”消息。

[0181] 表 31

[0182]

stereo_video_info(payloadSize) {	C	描述符
field_views_flag	5	u(1)
if(field_views_flag)		
top_field_is_left_view_flag	5	u(1)
else{		

current_frame_is_left_view_flag	5	u(1)
next_frame_is_second_view_flag	5	u(1)
}		
left_view_self_contained_flag	5	u(1)
right_view_self_contained_flag	5	u(1)
}		

[0183] “Field_views_flag”表示是否存在基于场的立体流。当“Top_field_is_left_view_flag”是“1”时,它表示以垂直方向交织格式(左侧视图优先)而形成的立体内容,而当“Top_field_is_left_view_flag”是“0”时,它表示以垂直方向行交织格式(右侧视图优先)而形成的立体内容。当“Current_frame_is_left_view_flag”是“1”时,它表示当前帧表现了左侧视图,而当它是“0”时,表示当前帧示出了右侧视图。当“Next_frame_is_second_view_flag”是“1”时,它表示立体图像由当前帧和下一帧形成,而当它是“0”时,表示立体图像由当前帧和前一帧形成。当“Left_view_self_contained_flag”是“1”时,它表示将流包覆(coat)为与右侧视图没有相关性的独立流,而当它是“0”时,表示基于与右侧视图的相关性来包覆流。当“Right_view_self_contained_flag”是“1”时,它表示将流包覆为与左侧视图没有相关性的独立流,而当它是“0”时,表示基于与左侧视图的相关性来包覆流。

[0184] “立体视频信息 SEI”信息包括“stereoscopic_composition_type”之中的表 32 的格式。然而,不支持“并排”格式

[0185] 表 32

[0186]

1	垂直方向交织格式
2	帧序格式
3	场序格式
4	单视场左侧图像
5	单视场右侧图像

[0187]

[0188] 在下文中,引入了用于使用典型的 AVC SEI 信息来服务立体内容的方法和存储格式。仅当通过 AVC 来编码立体内容时,它是可能的。

[0189] 引入使用“reserved_sei_message”的 SS-VAF,作为添加基于典型的“立体视频信息 SEI”的每个立体流所需的相机参数和显示信息。同样,可以像表 33 一样扩展和使用“立体视频信息 SEI”。当“Side_by_side_flag”是“1”时,在左侧视图优先的情况下,将它形成为一帧中的左侧图像和右侧图像。当“Side_by_side_flag”是“0”时,在右侧视图优先的情况下,将它形成为一帧中的左侧图像和右侧图像。在表 33 中,“C”表示语法的类别,而“u(1)”意指使用 1 比特的“无符号整数”。

[0190] 表 33

[0191]

stereo_video_info(payloadSize) {	C	描述符
field_views_flag	5	u(1)
if(field_views_flag)		
top_field_is_left_view_flag	5	u(1)
else{		
current_frame_is_left_view_flag	5	u(1)
next_frame_is_second_view_flag	5	u(1)
}		

else{		
left_view_self_contained_flag	5	u(1)
right_view_self_contained_flag	5	u(1)
}		
side_by_side_flag		u(1)
}		

[0192] 表 34 定义了使用 AVC 的 SEI 信息之中的“reserved_sei_message (有效负载尺寸)”的立体相机信息。这里,可以添加其他的相机信息。可以独立地使用所添加的信息。基于此,可以获得用于立体内容流的相机参数信息。

[0193] 表 34

[0194]

stereo_camera_info(payloadSize) { if(Is_CamParams){ baseline; focallength; convergence_distance; if(Is_camera_cross) { rotation[]; translation[]; } }	C 5 5 5 5 5	描述符 U(32) U(32) U(32) U(32) U(32)
--	------------------------------------	--

[0195] 表 35 定义了用于使用 AVC 的 SEI 信息之中的“reserved_sei_message (有效负载尺寸)”来显示立体内容的信息。基于在表 35 中定义的信息,提取立体内容视差值。

[0196] 表 35

[0197]

stereo_display_info(payloadSize) { if(Is_disInfo){ MinofDisparity; MaxofDisparity; }	C 5 5	描述符 U(16) U(16)
--	-----------------	---------------------------

[0198] 然而,可能通过将上面信息组合为一个 SEI_message (SEI 消息) 来提供立体相机和显示信息。

[0199] 图 13 是图示了使用立体视频信息和保留的 SEI 的 SS-VAF 的图。在所示的应用格式中选择性地包括 LAsER。

[0200] 在立体内容由 2D/3D 混合流形成的情况下,通过“iloc”逻辑框而在 3D 流片段期间获得在 AVC 流 SEI 消息中定义的 3D 流信息。这里,3D 流信息可以包括“立体视频信息 SEI”、“立体相机信息 SEI”、和“立体显示信息 SEI”。在立体内容仅仅由单一格式的 3D 流形成的情况下,可以使用在 AVC 流 SEI 消息中定义的 3D 流信息来表达立体内容。这里,3D 流信息可以包括“立体 视频信息 SEI”、“立体相机信息 SEI”、和“立体显示信息 SEI”。

[0201] 图 14 是图示了当立体内容包括两个基础流(ES)并且仅仅由单一格式立体流形成时的 SS-VAF 的图。在 ES=2 的情况下,在每个对应的“trak”中包括左侧流信息和右侧流信息。这里,必须示出左侧流和右侧流之间的关系。例如,如果左侧图像是主要图像而右侧图像是补充图像,则可能通过示出在包括右侧图像流信息的“trak”逻辑框和包括左侧图像流信息的另一“trak”逻辑框之间的关系来去除“ishd”信息的冗余。这种关系使用在基于 ISO 的文件格式中包括的“tref”逻辑框。它允许描述用于再现所必须的所有 trak_ID (轨道 ID)。因此,利用右侧图像流(补充图像)中的“trak”中的“tref”来描述所有的 trak_ID。

[0202] 表 36 示出了根据本发明实施例的支持各个立体配置格式和相机参数情况下的“ishd”逻辑框的语法。“item_ID”表示定义了下一信息的项目的 ID,并且具有大于 1 的值。当“current_indicator”是“1”时,它表示接下来描述的信息的有效性,而当它是“0”时,表示先前描述的信息与接下来描述的信息相同。然而,如果 item_ID=1,则它意指不存在接下来将描述的信息。即,终端基于“current_indicator”来确定接下来将描述的信息的有效性,并且当它是 0 时,确定出它与先前描述的“ishd”信息相同。

[0203] 同样,“LR_first”表示左侧图像和右侧图像的参考位置选择。“Is_camParams”表示是否存在相机参数。“Is_displaySafeInfo”表示是否存在立体内容显示信息。“Baseline”表示左侧相机和右侧相机之间的距离。“focallength”表示从 CCD 到图像平面(底片)的距离。当“Is_camera_cross”是“1”时,它表示交叉轴相机,当它是“0”时,表示平行轴相机。

[0204] 表 36

[0205]

```
Aligned(8) class Initial Stereoscopic Header box extend FullBox('ishd',  
version=0,0)
```

```
    unsigned int(16) item_ID;  
    unsigned int(1)  current_indicator;  
    unsigned int(8)  Stereoscopic_Composition_Type;  
    unsigned int(1)  LR_first;  
    unsigned int(1)  Is_camParams;
```

[0206]

```

unsigned int(1) Is_displaySafeInfo;
unsigned int(4) Reserved;

//所有以下字段是可选字段
if(Is_CamParams){
    unsigned int(32) baseline;
    unsigned int(32) focallength;
    unsigned int(1) Is_camera_cross
    if(Is_camera_cross) {
        unsigned int(32) convergence_distance;
        unsigned int(32) rotation[];
        unsigned int(32) translation[];
        unsigned int(7) reserved;
    }
    if(Is_displaySafeInfo){
        unsigned int(16) ViewingDisatance;
        int(16) VerticalDisparity;
        int(16) MinofDisparity;
        Int(16) MaxofDisparity;
    }
    //其他附加信息
    if(other_flag){}

    aligned(8) class Initial Stereoscopic Header box extend FullBox('ishd',
version=0,0){
        unsigned int(16) item_count;
        ishdEntry[item_count] ishds;
    }
}

```

[0207] 此外，“convergence_distance”表示从基线的中心到会聚点的距离(在平行相机的情况下，“convergence_distance”具有无穷大的值。当所有的比特都是1时，它表示无穷大的距离。)“rotation (旋转)”表示对于对象的相机方位角，而“translation (平移)”表示立体相机的运动(当所有的比特都是0时，它表示没有立体相机运动)。同样，“VerticalDisparity (垂直方向视差)”表示左侧和右侧图像的垂直方向视差尺寸，“MinofDisparity”表示左侧和右侧图像的最小水平方向视差尺寸，“MaxofDisparity”表示左侧和右侧图像的最小 视差尺寸，而“项目计数”表示下一阵列中的条目的数目。

[0208] 表 37 示出了根据本发明第一实施例的用于支持各个相机参数的“ishd”逻辑框的语法。这里，如果立体配置格式相同，则参考右侧前一“ishd”信息。同样，可以将“ishd”中包括的相机参数和显示信息分割到附加逻辑框，并进行存储。

[0209] 表 37

[0210]

```
Aligned(8) class Initial Stereoscopic Header box extend FullBox('ishd',
version=0,0)
    unsigned int(16) item_ID;
    unsigned int(8) Stereoscopic_Composition_Type;
    unsigned int(1) LR_first;
    unsigned int(1) current_indication;
    unsigned int(1) Is_camParams;
    unsigned int(1) Is_displaySafeInfo;
    unsigned int(4) Reserved;

    //所有以下字段是可选字段
    if(Is_CamParams){
        unsigned int(32) baseline;
        unsigned int(32) focallength;
        unsigned int(1) Is_camera_cross
        if(Is_camera_cross){
            unsigned int(32) convergence_distance;
            unsigned int(32) rotation[];
            unsigned int(32) translation[];
        }
        unsigned int(7) reserved;
    }
    if(Is_displaySafeInfo){
        unsigned int(16) ViewingDisatance;
        int(16) VerticalDisparity;
        int(16) MinofDisparity;
        int(16) MaxofDisparity;
    }
    //其他附加信息
```

[0211]

```

        if(other_flag){}

aligned(8) class Initial Stereoscopic Header box extend FullBox('ishd',
version=0,0){
    unsigned int(16) item_count;
    ishdEntry[item_count] ishds;
}

```

[0212] 表 38 示出了根据本发明第二实施例的用于支持各个相机参数的“ishd”逻辑框的语法。假设立体配置格式相同并且参考右侧前一“ishd”。

[0213] 表 38

[0214]

```

Aligned(8) class Initial Stereoscopic Header box extend FullBox('ishd',
version=0,0)

    unsigned int(8) Stereoscopic_Composition_Type;
    unsigned int(1) LR_first;
    unsigned int(7) Reserved;
    unsigned int(16) item_count;
    for(i=0; i<item_count;i++) {
        unsigned int(16) item_ID;
        unsigned int(1) current_indication;
        unsigned int(1) Is_camParams;
        unsigned int(1) Is_displaySafeInfo;
        unsigned int(5) Reserved;

//所有以下字段是可选字段
        if(Is_CamParams){
            unsigned int(32) baseline;
            unsigned int(32) focallength;
            unsigned int(1) Is_camera_cross
            unsigned int(7) reserved;
            if(Is_camera_cross){
                unsigned int(32) convergence_distance;
                unsigned int(32) rotation[];
                unsgiend int(32) translation[];
            }
        }
    }

```

[0215]

```

    }
  }
  if(Is_displaySafeInfo){
    unsigned int(16) ViewingDisatance;
    int(16) VerticalDisparity;
    int(16) MinofDisparity;
    int(16) MaxofDisparity;
  }
  //其他附加信息
  if(other_flag){}
}

```

[0216] 表 39 示出了根据本发明第三实施例的用于支持各个相机参数的“ishd”逻辑框的语法。假设立体配置格式相同并且参考诸如“cameParams（相机参数）”和“displaysafeInfo（显示安全信息）”之类的预定 Item_ID。

[0217] 在表 39 中,当“Is_ref”是“0”时,它表示不存在所参考的相机参数和显示信息,而当它是“1”时,表示存在所参考的 Item_ID。“current_refIndex”表示所参考的项目 ID。

[0218] 表 39

[0219]

```

Aligned(8) class Initial Stereoscopic Header box extend FullBox('ishd',
version=0,0)

```

```

    unsigned int(8)    Stereoscopic_Composition_Type;
    unsigned int(1)    LR_first;
    unsigned int(7)    Reserved;
    unsigned int(16)   item_count;
    for(i=0; i<item_count;i++) {
        unsigned int(16) item_ID;
        unsigned int(1) Is_camParams;
        unsigned int(1) Is_displaySafeInfo;
        unsigned int(1) Is_ref;
        unsigned int(5) Reserved;
    }

```

//所有以下字段是可选字段

[0220]

```

If(Is_ref){
    unsigned int(16)    current_refIndex;
}
else{
    if(Is_CamParams){
        unsigned int(32) baseline;
        unsigned int(32) focallength;
        unsigned int(1)  Is_camera_cross
        unsigned int(7)  reserved;
    }
    if(Is_camera_cross) {
        unsigned int(32) convergence_distance;
        unsigned int(32) rotation[];
        unsigned int(32) translation[];
    }
    if(Is_displaySafeInfo){
        unsigned int(16) ViewingDisatance;
        int(16)    VerticalDisparity;
        int(16)    MinofDisparity;
        int(16)    MaxofDisparity;
    }
}
//其他附加信息
if(other_flag){}
}

```

[0221] 表 40 示出了根据本发明第四实施例的用于支持各个相机参数的“ishd”逻辑框的语法。假设立体配置格式相同并且“cameParams”和“displaysafeInfo”参考不同的 Item_ID。

[0222] 表 40

[0223]

```

Aligned(8) class Initial Stereoscopic Header box extend FullBox('ishd',
version=0,0)

    unsigned int(8) Stereoscopic_Composition_Type;
    unsigned int(1) LR_first;

```

[0224]

```

unsigned int(7) Reserved;
unsigned int(16)    item_count;
for(i=0; i<item_count;i++) {
    unsigned int(16) item_ID;
    unsigned int(1)  Is_camParams;
    unsigned int(1)  Is_displaySafeInfo;
    unsigned int(1)  Is_camParamsref;
    unsigned int(1)  Is_displaySafeInforef;
    unsigned int(4)  Reserved;

    //所有以下字段是可选字段
    If(Is_camParamsref || Is_displaySafeInforef){
        unsigned int(16) current_camrefIndex;
        unsigned int(16) current_displayrefIndex;
    }
    else{
        if(Is_CamParams) {
            unsigned int(32) baseline;
            unsigned int(32) focallength;
            unsigned int(1)  Is_camera_cross
            unsigned int(7)  reserved;
            if(Is_camera_cross) {
                unsigned int(32) convergence_distance;
                unsigned int(32) rotation[];
                unsgiend int(32) translation[];
            }
            if(Is_displaySafeInfo){
                unsigned int(16) ViewingDisatance;
                int(16)  VerticalDisparity;
                int(16)  MinofDisparity;
                int(16)  MaxofDisparity;
            }
        }
        //其他附加信息
        if(other_flag){}
    }
}

```

[0225] 在表 40 中,当“Is_camParamsref”是“0”时,它表示不存在所参考的相机参数信息,而当它是“1”时,表示存在所参考的 Item_ID。当“Is_displaySafeInforef”是“0”

时,它表示不存在所参考的显示安全信息,而当它是“1”时,表示存在所参考的 Item_ID。
“current_cameraIndex”表示所参考的 Item_ID,而“Current_displayIndex”表示所参考的 Item_ID。

[0226] 可以将“ishd”逻辑框划分为用于记录立体视频媒体信息的“svmi”逻辑框(立体视频媒体信息逻辑框)和用于记录相机参数和显示信息的“scdi”逻辑框(立体相机和显示信息逻辑框)。由于“svmi”逻辑框是强制性的而“scdi”逻辑框不是强制性的,所以有利地将“ishd”逻辑框划分为“svmi”逻辑框和“scdi”逻辑框,以便去除不必要的信息。

[0227] “svmi”逻辑框提供立体视觉类型和片段信息。更详细地,立体视频媒体信息包括关于立体图像配置类型的信息、关于左侧图像和右侧图像之间首先编码的图像的信息、当立体内容的基础流从立体片段改变为单视场片段或者从单视场片段改变为立体片段时的关于片段的数目的信息、关于连续采样的数目或者计数开始采样数目的信息、以及关于当前采样是否是立体的信息。

[0228] “scdi”逻辑框包括关于是否存在相机参数的信息、左侧相机和右侧相机之间的距离、左侧和右侧相机的安排、从主要视图相机到子视图相机的相对角、以及在左侧图像和右侧图像之间的最大视差和最小视差。表 41 示出了根据本发明实施例的“scdi”逻辑框的语法。

[0229] 表 41

[0230]

```
unsigned int(16) item_count;
for(i=0; i<item_count;i++) {
    unsigned int(16) item_ID;
    unsigned int(1) Is_camParams;
    unsigned int(1) Is_displaySafeInfo;
    unsigned int(1) Is_ref;
    unsigned int(5) Reserved;

    //所有以下字段是可选字段
    If(Is_ref){
        unsigned int(16) current_refIndex;
    }
```

[0231]

```

else{
    if(Is_CamParams){
        unsigned int(32) baseline;
        unsigned int(32) focallength;
        unsigned int(1) Is_camera_cross
        unsigned int(7) reserved;
    }
    if(Is_camera_cross) {
        unsigned int(32) convergence_distance;
        unsigned int(32) rotation[];
        unsigned int(32) translation[];
    }
    if(Is_displaySafeInfo){
        unsigned int(16) ViewingDisatance;
        int(16) VerticalDisparity;
        int(16) MinofDisparity;
        int(16) MaxofDisparity;
    }
}
}

```

[0232] 如所示的,可能通过“tref”逻辑框(轨道参考逻辑框)来去除在每个轨道中包括的“scdi”信息的冗余。当ES=2时,每个轨道的“iloc”逻辑框划分立体片段,以便提供“scdi”信息。这里,每个轨道的立体片段具有相同的item_ID以及相同的相机参数和显示信息。通过“tref”的“svdp”来将基础流划分为主要/补充轨道。尽管仅仅在一个轨道中包括“iloc”逻辑框,但是当执行3D显示时,可能经由通过立体片段而同步“iloc”逻辑框来再现它。

[0233] 还可能通过“tref”逻辑框来去除在每个轨道中包括的相同立体视频媒体信息(“svmi”)的冗余。在通过“ftyp”逻辑框来识别立体内容之后,通过“tref”逻辑框的“svdp”逻辑框来划分主要/补充轨道。如果一个轨道包括“svmi”逻辑框,则可以自动地识别另一轨道。由于“svmi”逻辑框是强制性的,所以可以在主要/补充轨道中包括它。可以仅仅在主要轨道中包括“svmi”逻辑框。

[0234] 图15是图示了根据本发明实施例的当ES=2时的SS-VAF的图。如图15所示,SS-VAF包括“svmi”逻辑框和“scdi”逻辑框。

[0235] 当立体内容包括两个基础流(ES)时,包括两个轨道(“trak”)。可以将它划分为主要轨道和补充轨道。因此,使用补充轨道中的“tref”的“svdp”来参考主要轨道,并且提供在相关的“scdi”信息中包括的信息。这种结构具有去除在每个轨道中包括的相同“scdi”信息的冗余的优点。这里,track_ID表示所参考的轨道的ID。如果reference_type是“svdp”,则它还表示轨道包括用于参考轨道的立体相机和显示信息。

[0236] 当在用户正在利用3D显示模式来观看预定的图像的时候、用户将3D显示模式改

变为 2D 显示模式时,通过在终端处显示与主要轨道对应的图像来满足 2D 显示模式。其间,相同的单视场数据可以存在于具有两个基础流的轨道中的基础流的中间。在单视场数据中,在两个轨道中存储相同的内容。因此,单视场数据是不能显示为 3D 的数据。在此情况下,终端必须决定两个轨道之一,以显示其单视场数据。终端显示与根据本实施例而划分的主要轨道对应的图像。

[0237] 当立体视频的基础流是两个时,存在两个轨道。通过诸如传统的 LAsER 之类的场景描述符来将立体视频识别为两个对象,并且将立体视频连接到 LAsER 中的两个视频相关节点。然而,最终,必须将立体视频转换为一个 3D 视频格式,并且在终端中进行再现。因此,必须通过 LAsER 来将它识别为一个对象。即,由于需要将立体视频转换为一个 3D 视频格式,以用于再现立体视频的场景,所以它与所使用的一个节点相连。根据本实施例,使用“tref”中的“svdp”来将立体视频划分为主要 / 补充轨道,并且 LAsER 中的视频相关节点仅仅链接与主要轨道或媒体流对应的“轨道 ID”。

[0238] 图 16 是图示了根据本发明实施例的用于将立体存储为基于 ISO 的媒体文件格式的方法的流程图。首先,在步骤 S1602 中,在基于 ISO 的媒体文件格式的“mdat”逻辑框中存储目标立体内容。然后,在步骤 S1604 和 S1606 中,作为立体内容的元数据来在“moov”逻辑框中存储立体内容的立体视频媒体信息以及立体相机和显示信息。

[0239] 图 17 是图示了根据本发明另一实施例的用于以基于 ISO 的媒体文件格式来存储立体内容的方法的流程图。首先,在步骤 S1702 中,在基于 ISO 的媒体文件格式的“mdat”逻辑框中存储目标立体内容。然后,如果立体内容包括两个基础流,则在步骤 S1704 中在基于 ISO 的媒体文件格式的“tref”逻辑框中存储用于将基础流划分为主要轨道和补充轨道的信息(“svdp”)。然后,在步骤 S1706 中,存储用于立体内容的、具有仅仅链接到主要轨道的视频相关节点的 LAsER。这里,不使用链接到 LAsER 的部分,而仅仅使用“tref”。

[0240] 图 18 是图示了根据本发明又一实施例的用于存储立体内容的方法的流程图。这里,使用如上所述的 AVC SEI 来存储立体内容。首先,在步骤 S1802 中存储通过 AVC 而编码的立体内容。然后,在步骤 S1804 中,使用“reserved_sei_message”来存储对于立体内容的每个流所需的相机参数和显示信息。

[0241] 这里,“立体视频信息 SEI”附加地包括表示了立体图像配置是“并排”类型的信息。相机参数包括:左侧相机和右侧相机之间的距离、左侧和右侧一致的焦距(focal_length)、从基线到会聚点的距离、当左侧和右侧相机彼此交叉时左侧和右侧相机的旋转和平移中的至少一个。显示信息包括左侧图像和右侧图像之间的最大视差和最小视差。

[0242] 发明模式

[0243] 如上所述,可以将本发明的技术实现为程序,并且存储在诸如 CD-ROM、RAM、ROM、软盘、硬盘、和磁光盘之类的计算机可读记录介质中。由于本发明领域的技术人员可以容易地实现该处理,所以在这里将不提供进一步的描述。

[0244] 尽管已经针对具体实施例而描述了本发明,但是对于本领域技术人员明显的是,可以做出各种改变和修改,而不脱离由以下权利要求限定的本发明的精神和范围。

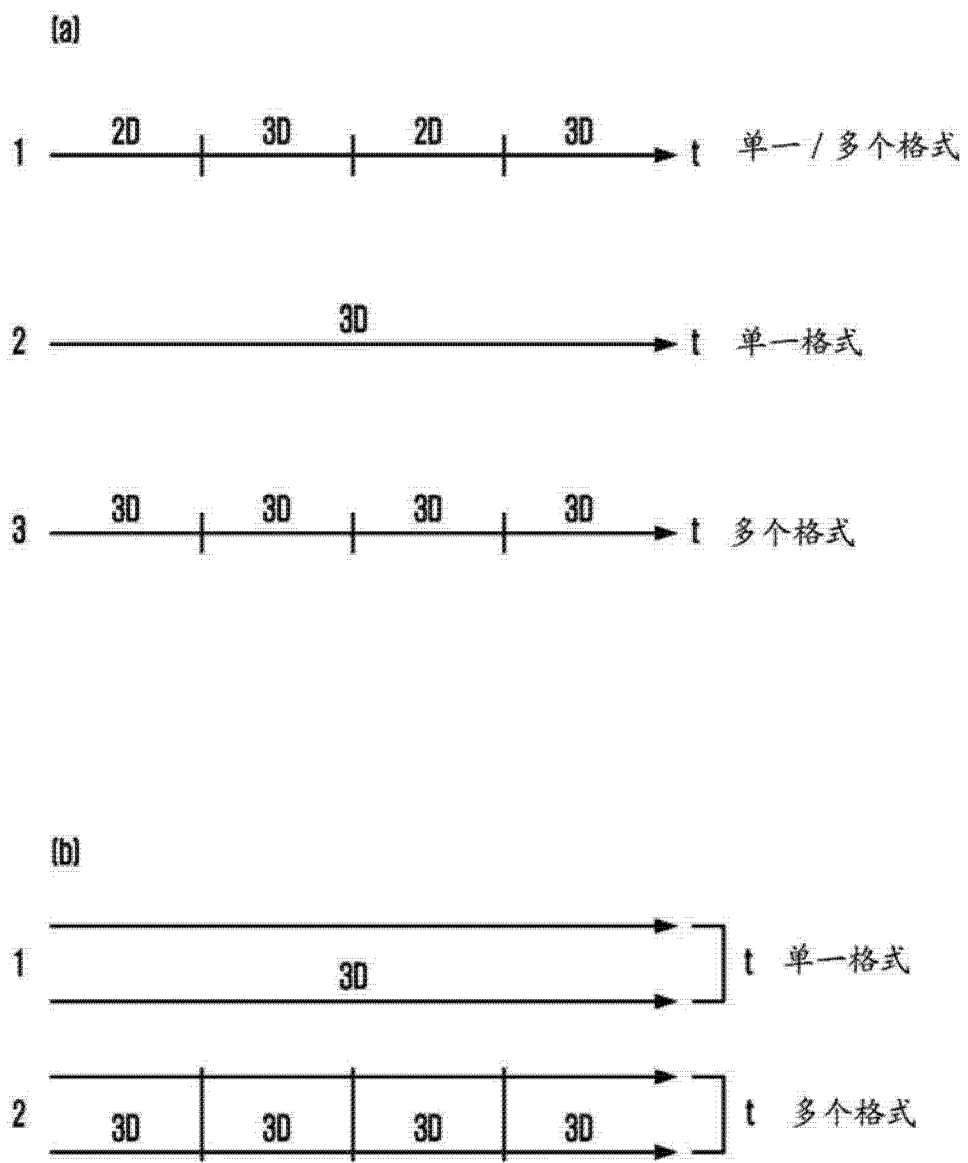


图 1

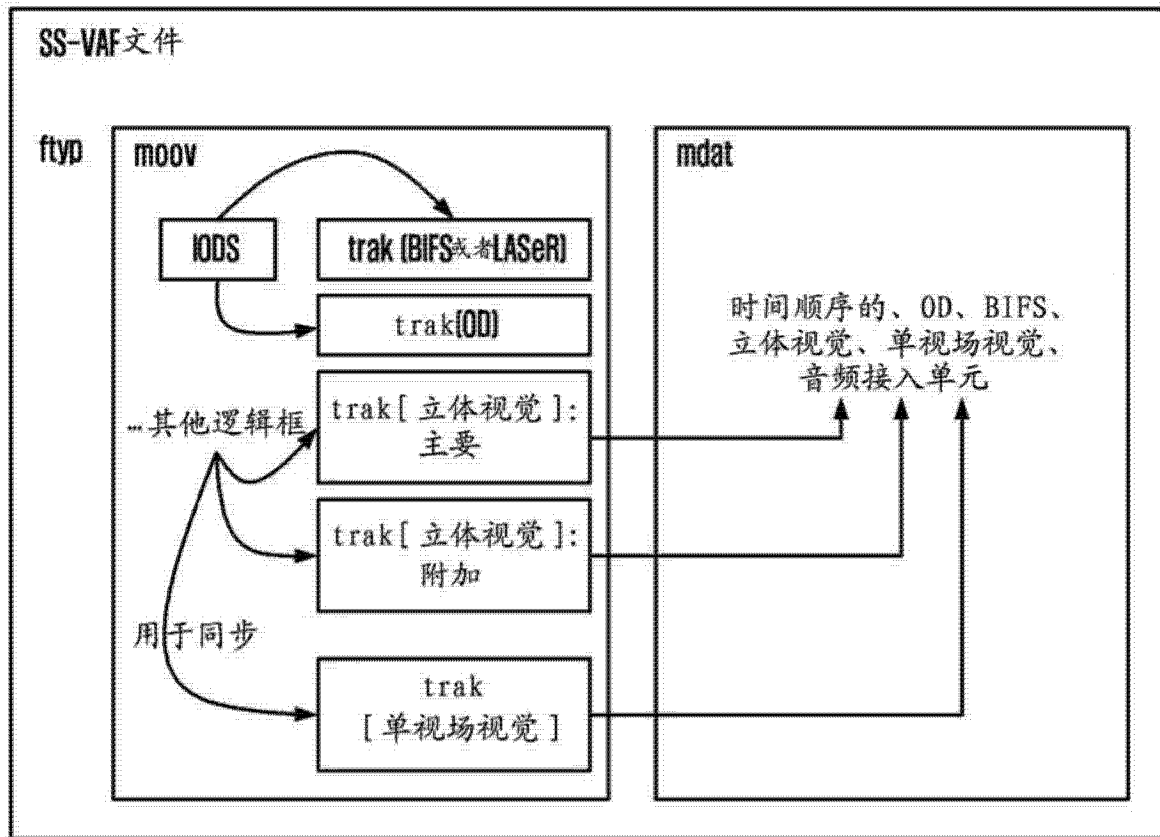


图 2

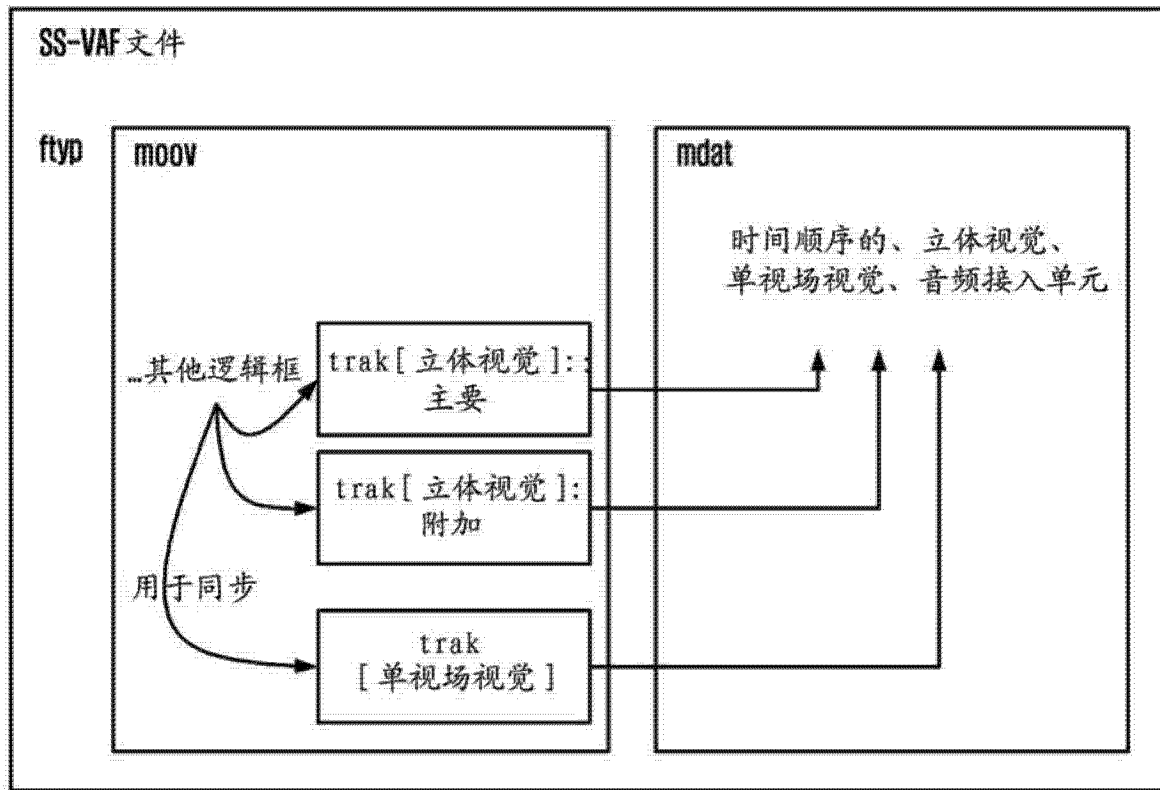


图 3

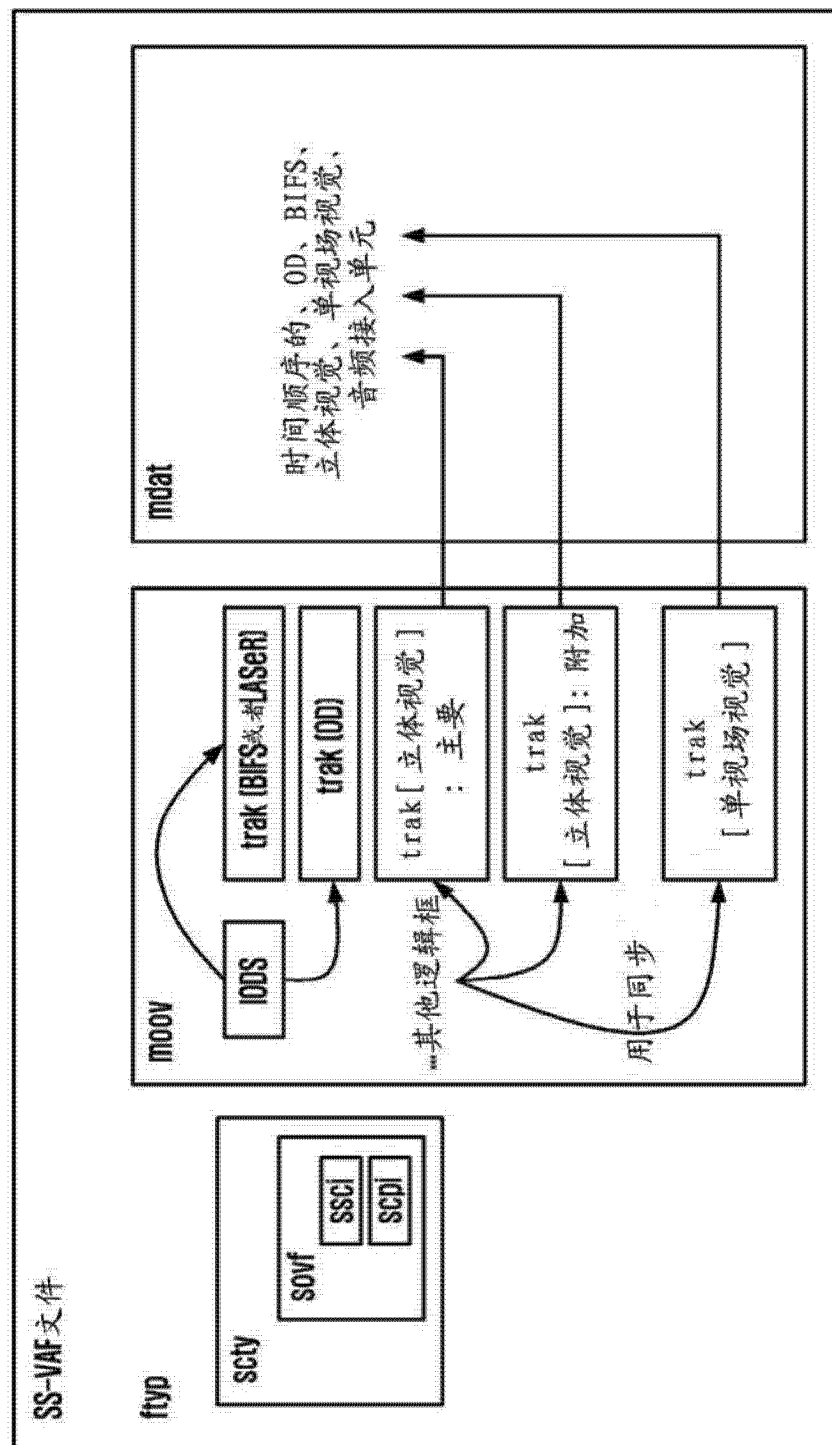


图 4

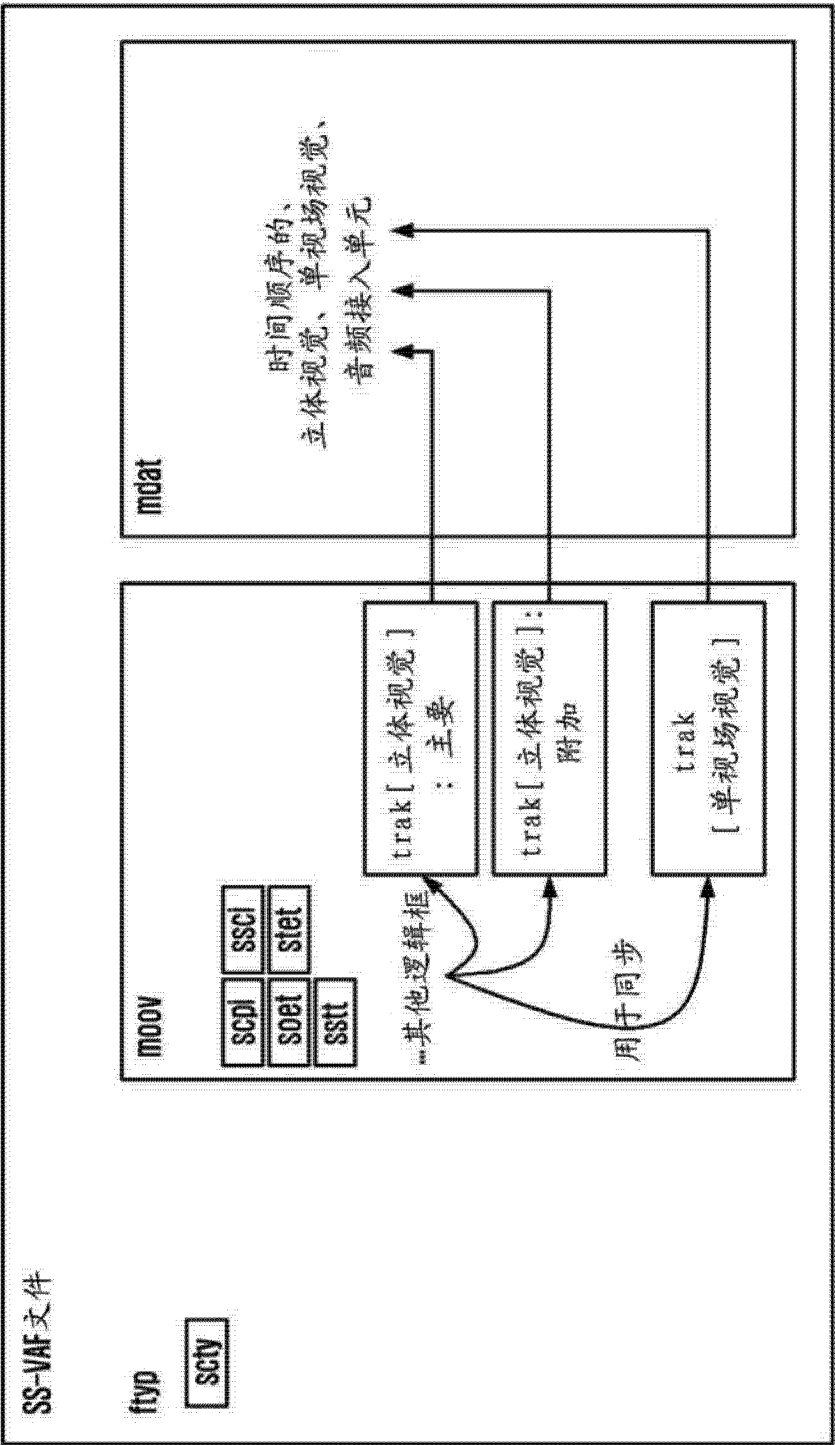


图 5

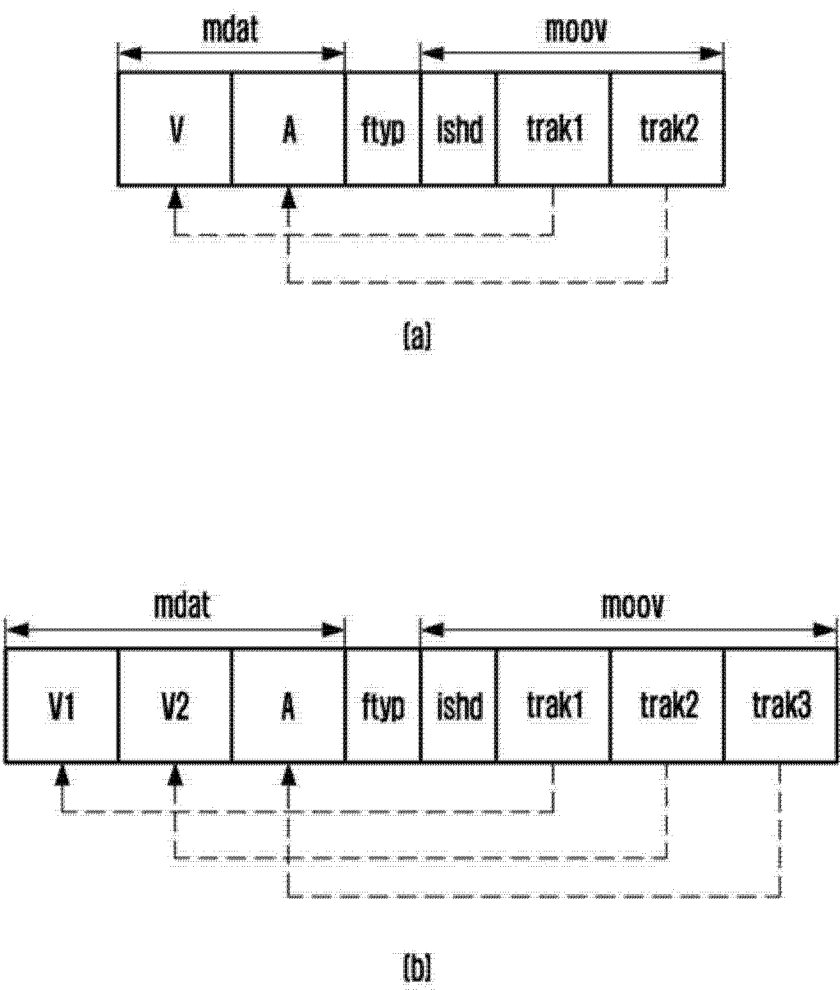


图 6

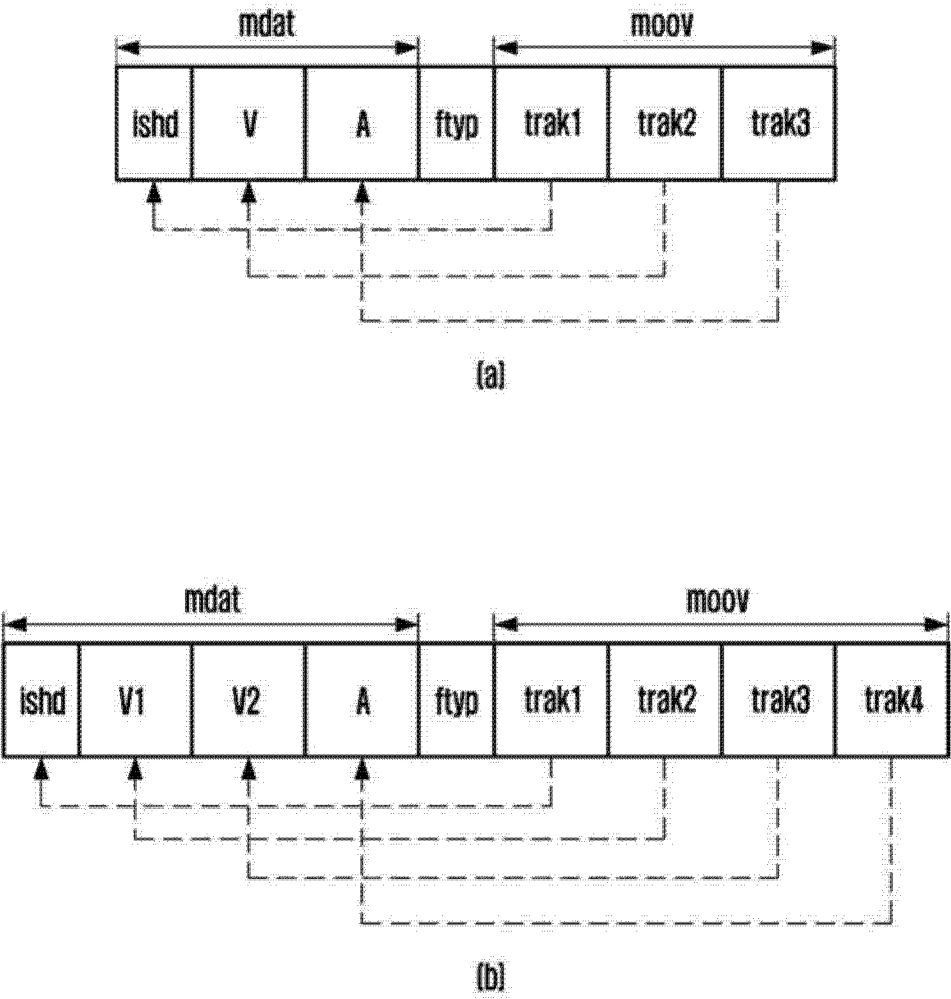


图 7

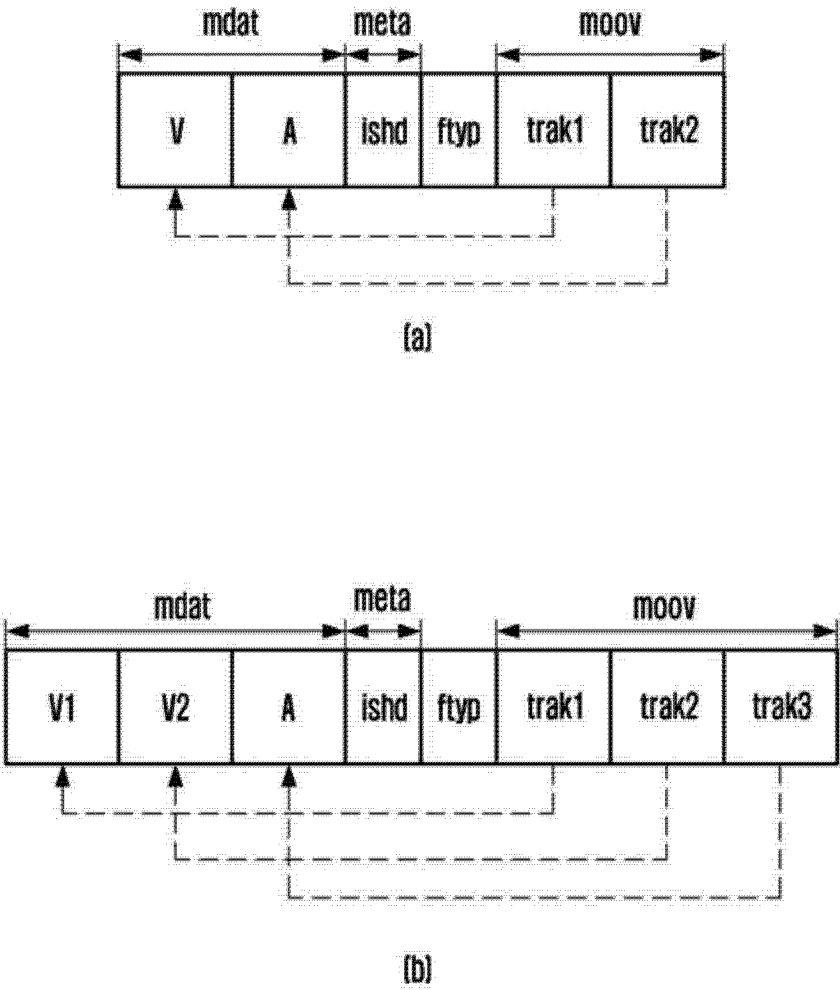


图 8

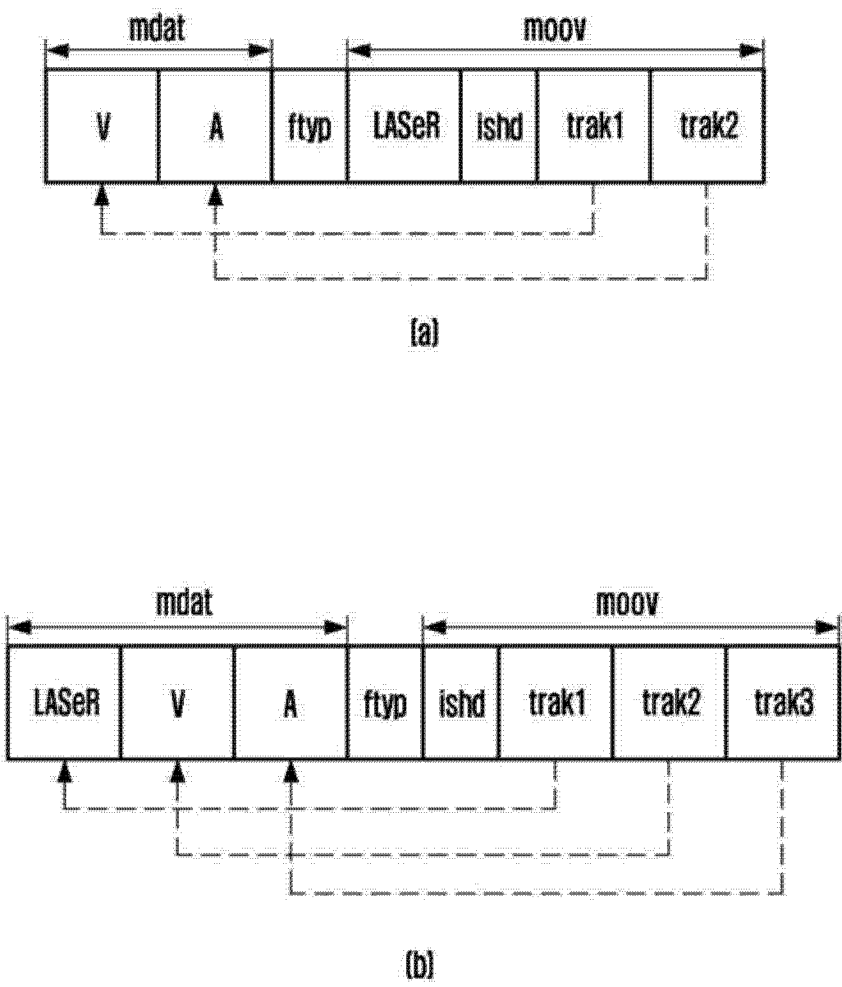


图 9

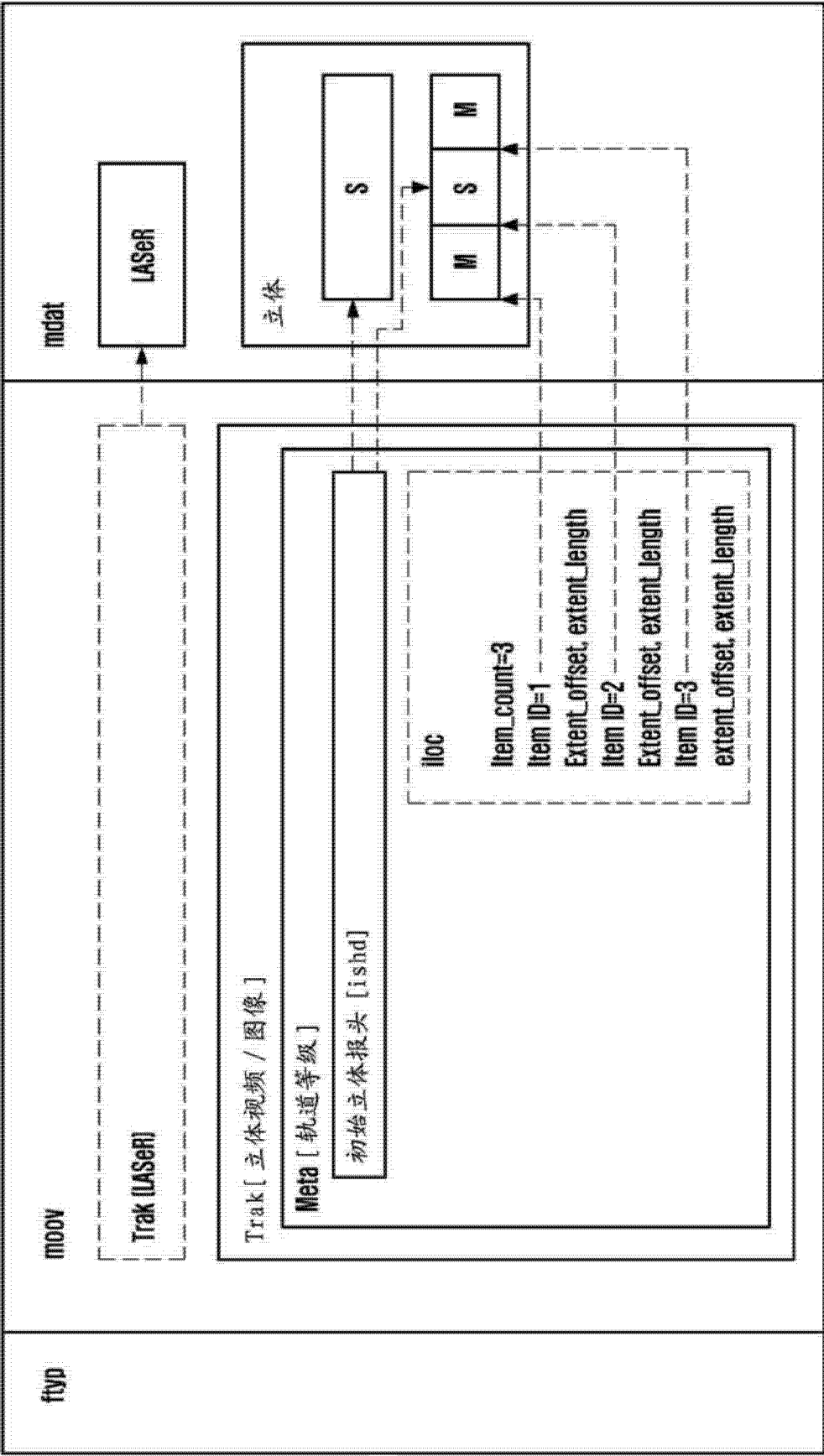


图 10

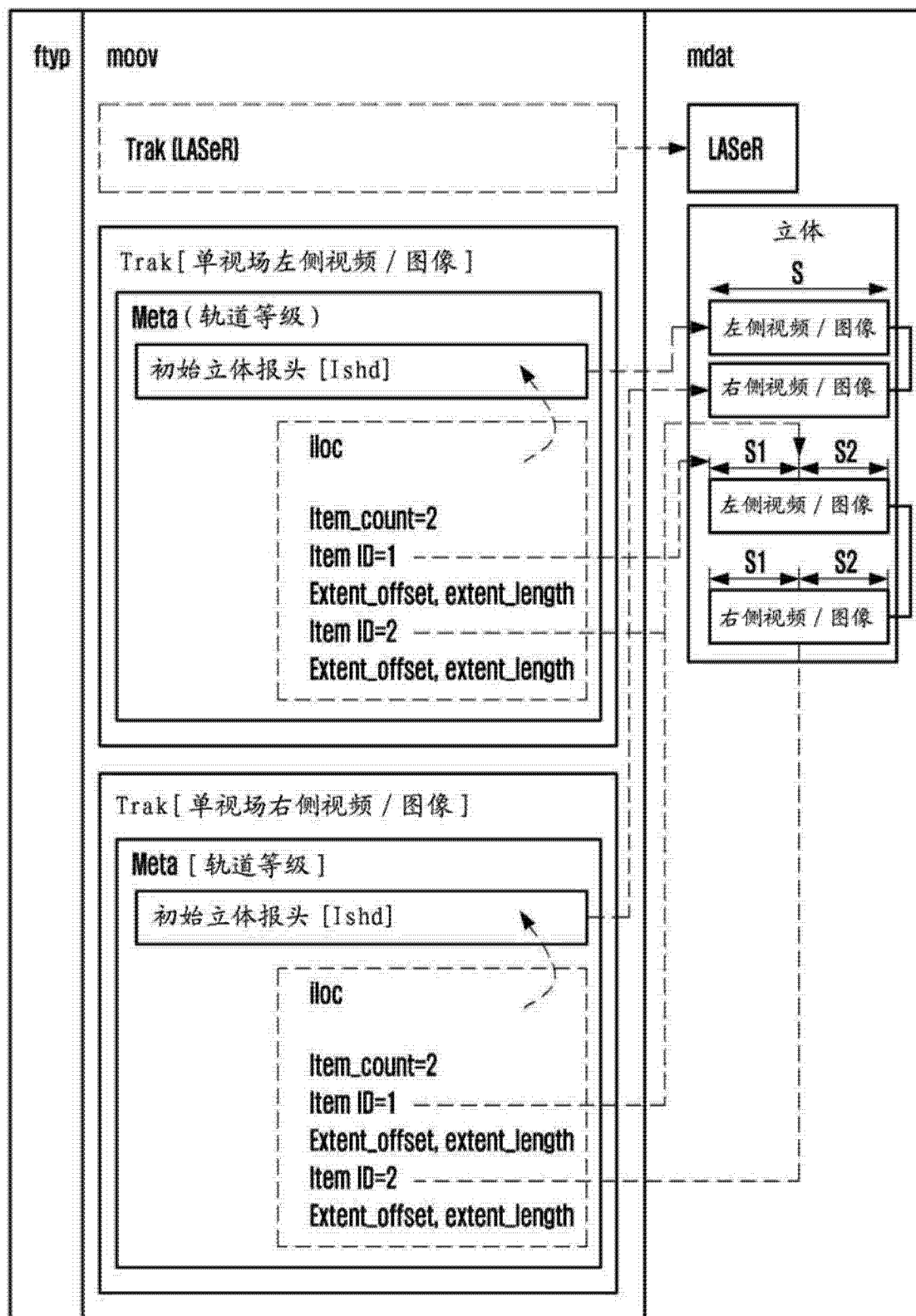
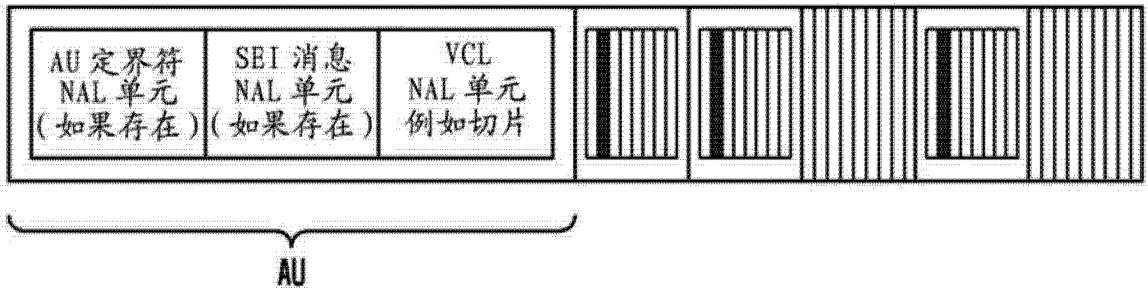


图 11

else if (payloadType == 21)		
stereo_video_info (payloadSize)	5	
else		
reserved_sei_message (payloadSize)	5	

[a]



[b]

图 12

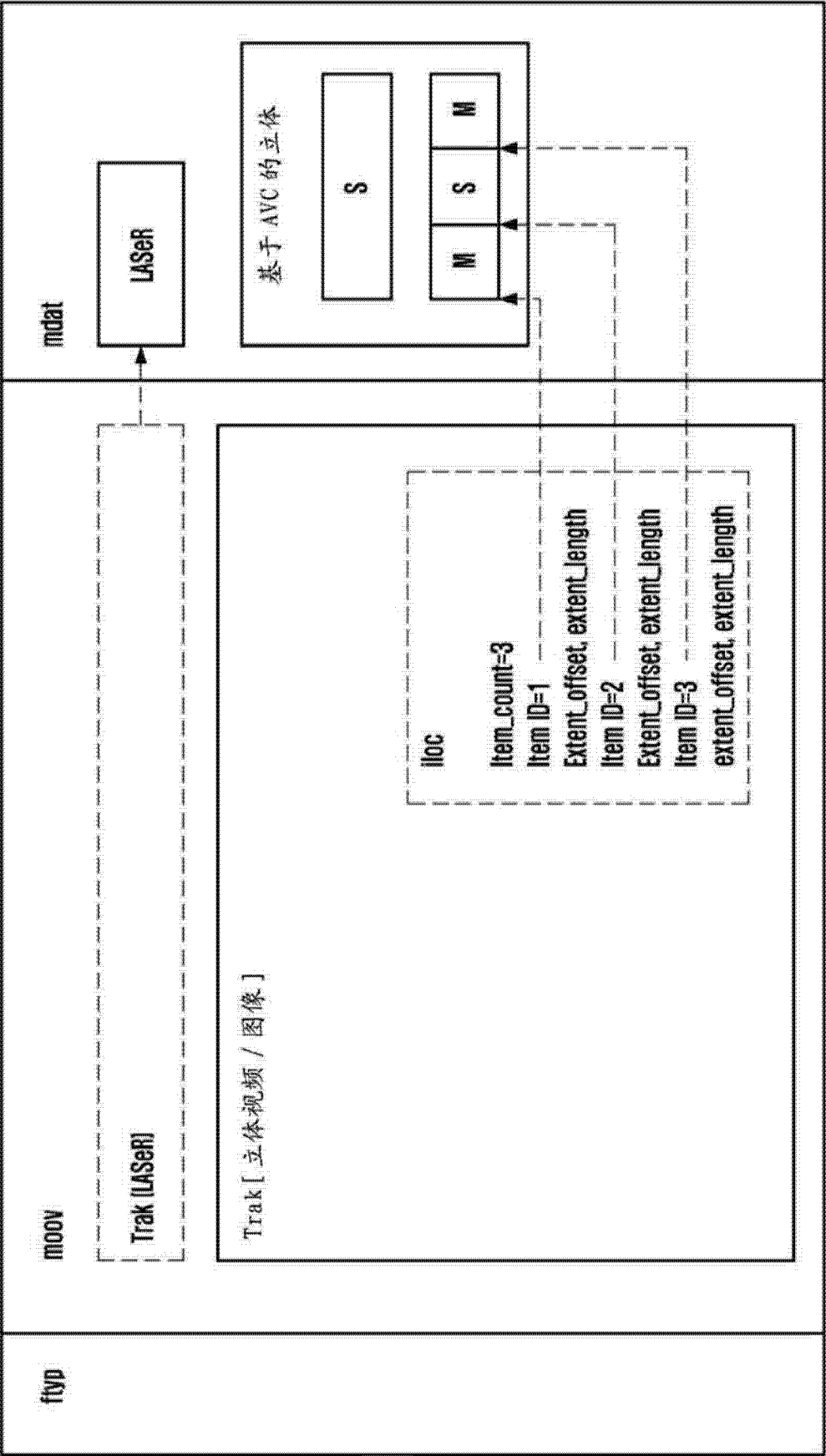


图 13

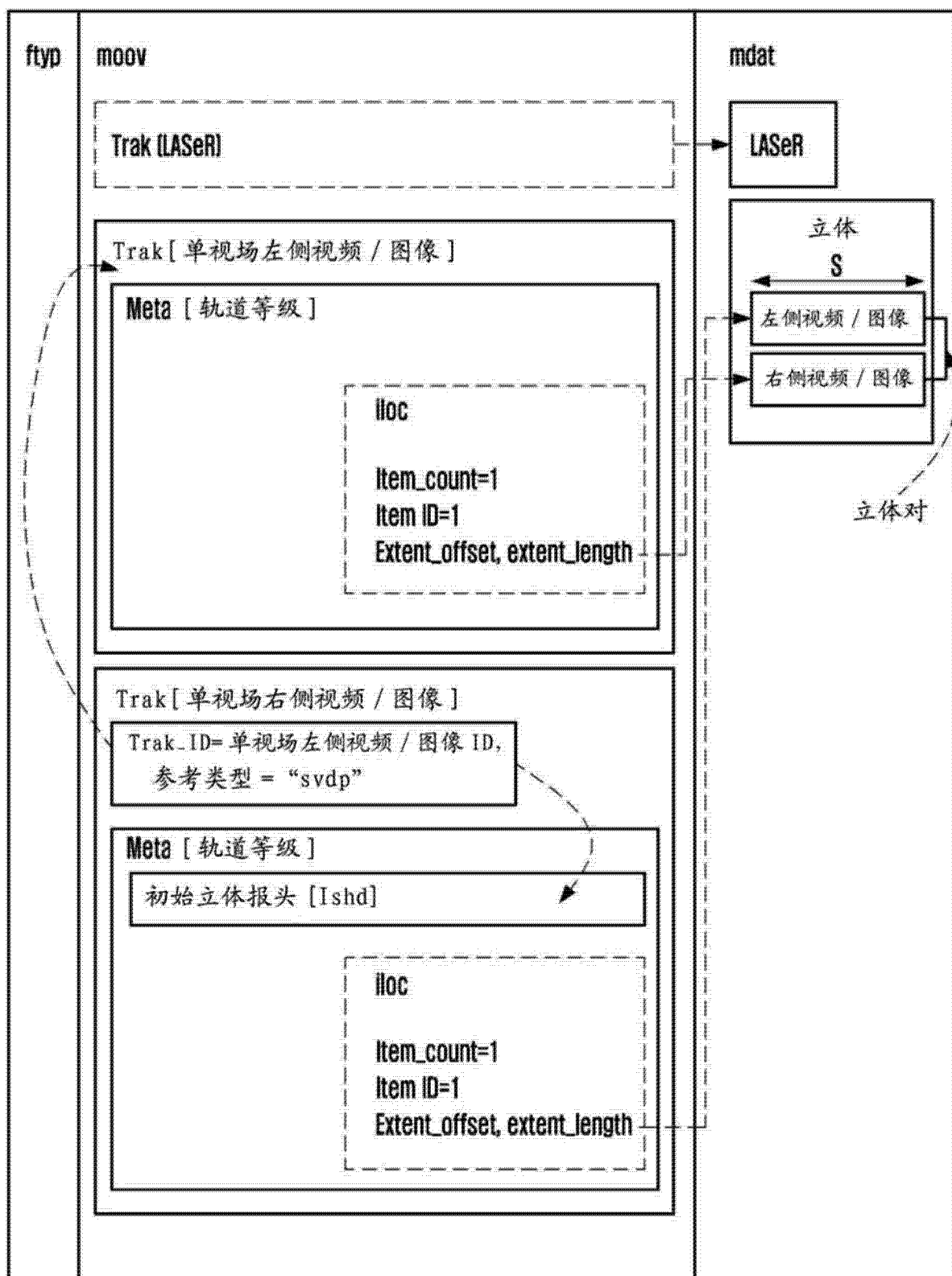


图 14

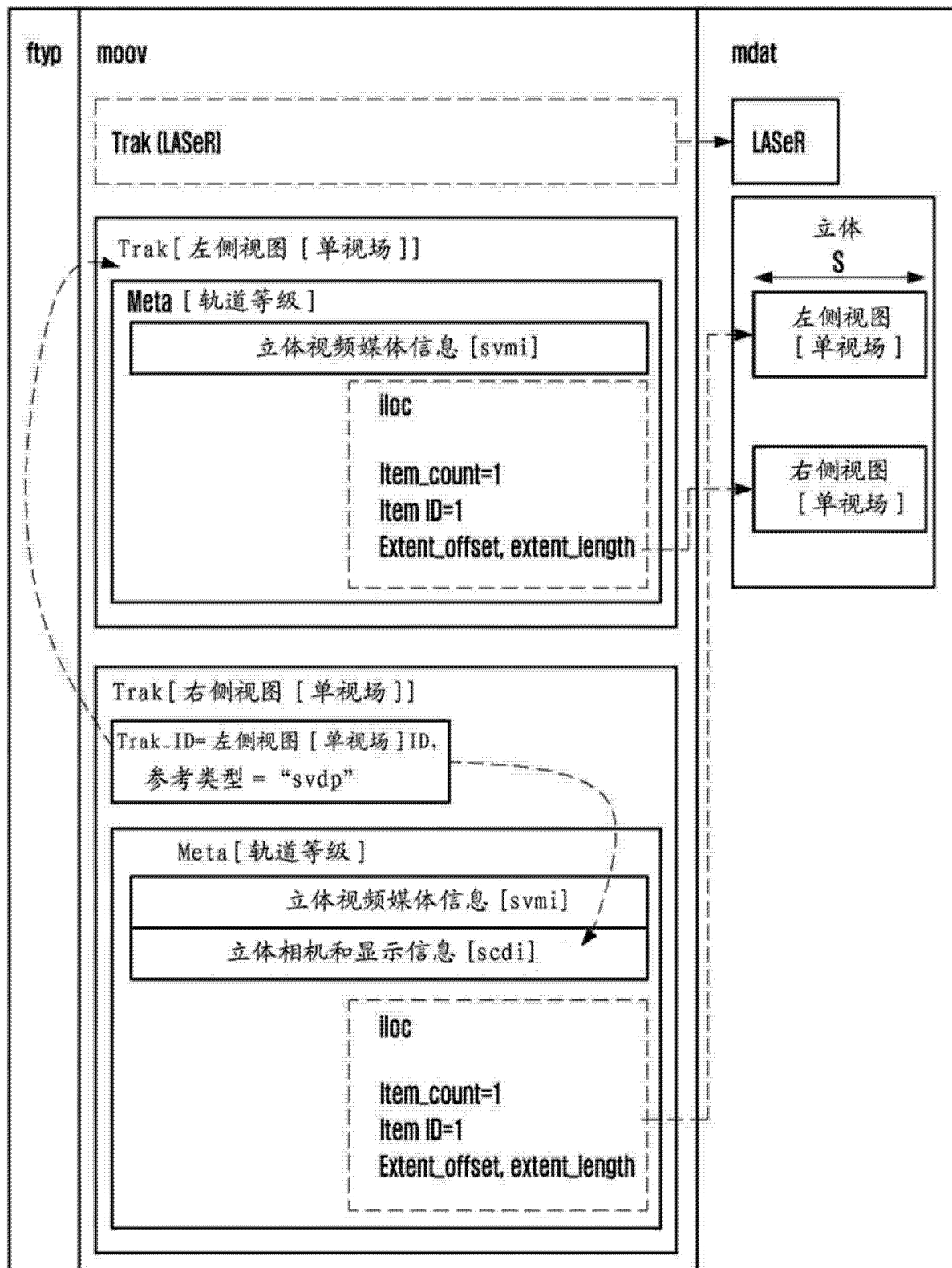


图 15

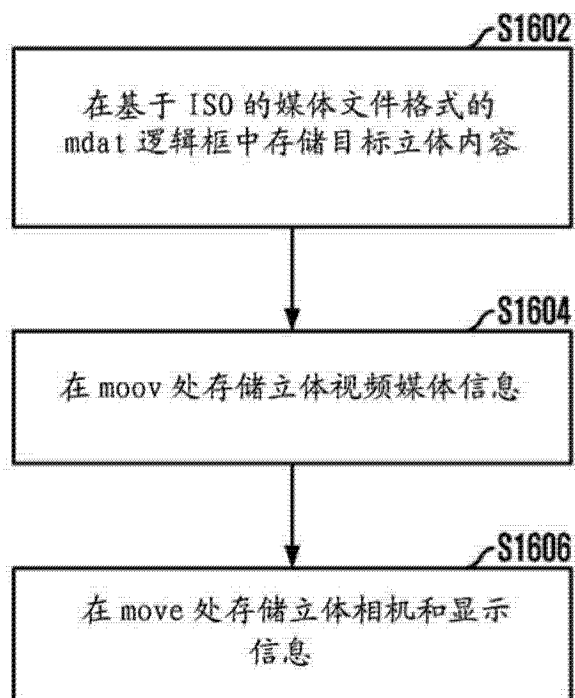


图 16

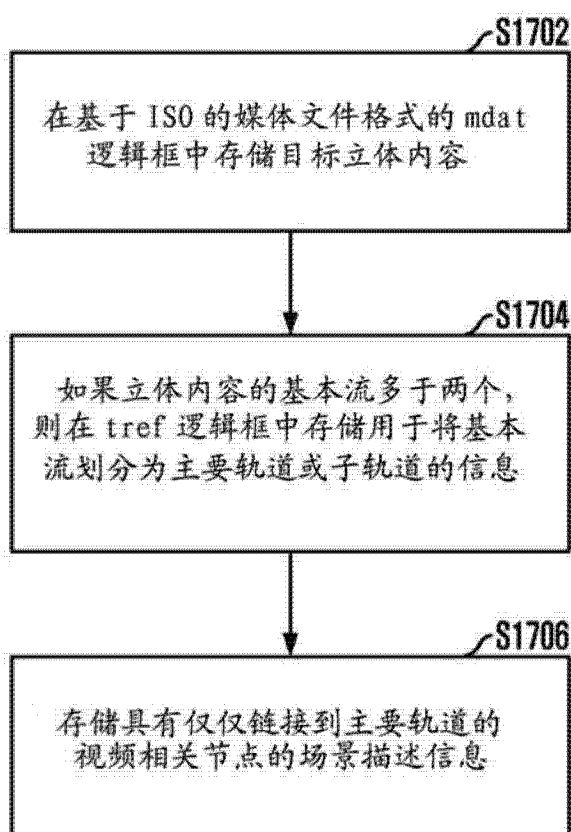


图 17

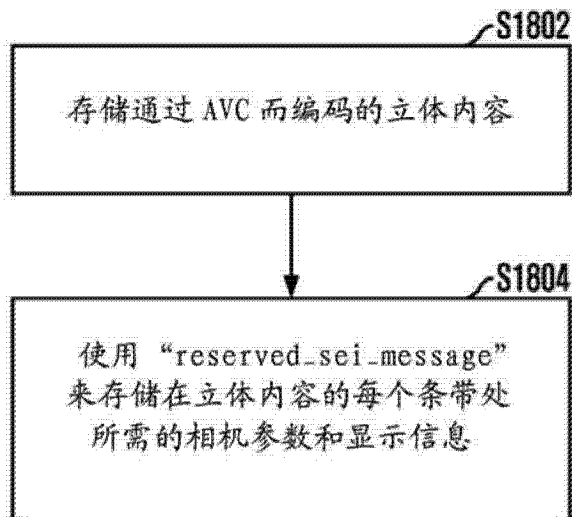


图 18