

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.



[12] 发明专利说明书

专利号 ZL 02827912.3

G06F 15/16 (2006.01)
G06F 9/00 (2006.01)
H04L 12/28 (2006.01)
H04J 3/24 (2006.01)

[45] 授权公告日 2007 年 10 月 10 日

[11] 授权公告号 CN 100342363C

[22] 申请日 2002.12.10 [21] 申请号 02827912.3

[30] 优先权

[32] 2001.12.12 [33] US [31] 60/340,388

[32] 2002.10.29 [33] US [31] 10/283,738

[86] 国际申请 PCT/US2002/039395 2002.12.10

[87] 国际公布 WO2003/050703 英 2003.6.19

[85] 进入国家阶段日期 2004.8.6

[73] 专利权人 索尼电子有限公司

地址 美国新泽西州

[72] 发明人 A·J·塔巴塔拜 T·沃尔克

M·Z·维沙拉姆

[56] 参考文献

US6240097B1 2001.5.29

US5953506A 1999.9.14

US6233253B1 2001.5.15

审查员 李 菲

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 杨 凯 王 勇

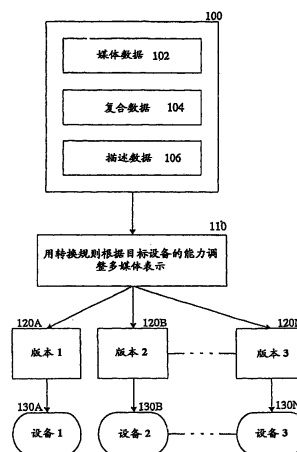
权利要求书 7 页 说明书 18 页 附图 8 页

[54] 发明名称

对多媒体数据进行转换以将其传送到多个异种设备

[57] 摘要

转换多媒体表示(100)，以在多个异种目标设备(130)上播放。根据目标设备(110)的能力选择转换操作，并将其用于根据多媒体表示的源版本创建该多媒体表示的适配版本(120)。该适配版本包含对应于该多媒体表示的媒体数据(102)的源版本的适配的媒体数据。



1. 一种用于转换多媒体数据以传送到多个目标设备的方法，包括：

根据目标设备（110）的能力从多个转换操作中选择转换操作；
以及

利用所述选择的转换操作并基于多媒体表示内的媒体数据的描述数据和复合数据，根据所述多媒体表示的源版本（100）创建适合于所述目标设备的所述多媒体表示的适配版本（120），所述多媒体表示的所述适配版本包含对应于所述多媒体表示的媒体数据（102）的源版本的适配的媒体数据。

2. 权利要求1的方法，其特征在于，创建适配版本包括：

将所述多媒体表示的描述数据的源版本转换为描述数据的目标版本（212）；以及

根据所述描述数据的所述目标版本生成所述适配的媒体数据（216）。

3. 权利要求1的方法，其特征在于，创建适配版本包括：

根据所述多媒体表示的描述数据的源版本导出所述媒体数据的源版本（208）；以及

将所述媒体数据的所述源版本转换为所述适配的媒体数据（210）。

4. 权利要求1的方法，其特征在于，创建适配版本包括：

为所述媒体数据的所述源版本中的多个媒体对象中的每个媒体对象准备适配的媒体对象。

5. 权利要求1的方法，其特征在于，创建适配版本包括：

如果所述媒体数据的所述源版本包括视频数据和图像数据其中至少之一项，则调整空间分辨率和时间分辨率其中至少一项。

6. 权利要求1的方法，其特征在于，创建适配版本包括：

根据期望的比特率调整所述媒体数据的所述源版本的比特率。

7. 如权利要求 6 的方法，其特征在于，所述期望的比特率基于用户偏好、传输媒体带宽和目标设备能力中的至少一项。

8. 权利要求 1 的方法，其特征在于，创建适配版本包括：生成所述媒体数据的概要形式的所述源版本。

9. 如权利要求 1 所述的方法，其特征在于，所述多媒体表示的所述适配版本还包括对应于所述多媒体表示的复合数据（420）的源版本的适配的复合数据（440）。

10. 权利要求 9 的方法，其特征在于，创建适配版本包括：

根据所述目标设备的能力和所述适配的媒体数据的特性，生成所述适配的复合数据。

11. 权利要求 9 的方法，其特征在于，创建适配版本包括：

将所述多媒体表示的描述数据的源版本转换为所述描述数据的目标版本（212）；以及

根据所述描述数据的所述目标版本生成所述适配的复合数据（216）。

12. 权利要求 9 的方法，其特征在于，创建适配版本包括：

根据所述多媒体表示的描述数据的源版本导出所述复合数据的所述源版本（208）；以及

将所述复合数据的所述源版本转换为所述适配的复合数据（210）。

13. 如权利要求 9 所述的方法，其特征在于，所述适配的复合数据包括所述适配的媒体数据中的多个媒体对象的空间和时间布局以及同步信息。

14. 如权利要求 9 所述的方法，其特征在于，所述多媒体表示的所述源版本还包括所述复合数据的所述源版本。

15. 如权利要求 1 所述的方法，其特征在于，选择转换操作包括将所选转换操作排序以满足最优化标准。

16. 如权利要求 1 所述的方法, 其特征在于, 所述转换操作是根据一组规则来选择的。

17. 如权利要求 1 所述的方法, 其特征在于, 所述目标设备的能力包括用于将所述适配的多媒体表示传送到所述目标设备的媒体的特性。

18. 如权利要求 1 所述的方法, 其特征在于, 所述转换操作还基于用户偏好。

19. 如权利要求 1 所述的方法, 其特征在于还包括:

将所述多媒体表示的所述适配版本传送到所述目标设备 (220)。

20. 如权利要求 1 所述的方法, 其特征在于还包括:

接收所述多媒体表示的所述源版本的媒体数据源版本、复合数据源版本和描述数据源版本中的至少一项 (200、202、206)。

21. 一种用于转换多媒体数据以传送到多个目标设备的系统, 包括:

通过总线 (740) 连接到存储器 (724) 的处理器 (722);

由所述处理器从所述存储器执行的转换过程, 所述转换过程使所述处理器根据目标设备 (110) 的能力从多个转换操作中选择转换操作, 利用所述选择的转换操作并基于多媒体表示内的媒体数据的描述数据和复合数据, 根据多媒体表示的源版本 (100) 创建适合于所述目标设备的所述多媒体表示的适配版本 (120), 所述多媒体表示的适配版本包含对应于所述多媒体表示的媒体数据 (102) 的源版本的适配的媒体数据。

22. 如权利要求 21 所述的系统, 其特征在于, 所述转换过程还使所述处理器在创建适配版本时, 将所述多媒体表示的描述数据的源版本转换为所述描述数据的目标版本 (212), 并根据所述描述数据的所述目标版本生成所述适配的媒体数据 (216)。

23. 如权利要求 21 所述的系统, 其特征在于, 所述转换过程还使所述处理器在创建适配版本时, 根据所述多媒体表示的描述数据的

源版本导出所述媒体数据的源版本（208），并将所述媒体数据的所述源版本转换为所述适配的媒体数据（210）。

24. 如权利要求 21 所述的系统，其特征在于，所述转换过程还使所述处理器在创建适配版本时，为所述媒体数据的所述源版本中的多个媒体对象中的每个媒体对象准备适配的媒体对象。

25. 如权利要求 21 所述的系统，其特征在于，所述转换过程还使所述处理器在创建适配版本时，如果所述媒体数据的所述源版本包括视频数据和图像数据中的至少一项，则调整空间分辨率和时间分辨率中的至少一项。

26. 如权利要求 21 所述的系统，其特征在于，所述转换过程还使所述处理器在创建适配版本时，根据期望的比特率调整所述媒体数据的所述源版本的比特率。

27. 如权利要求 26 所述的系统，其特征在于，所述期望的比特率基于用户偏好、传输媒体带宽和目标设备能力中的至少一项。

28. 如权利要求 21 所述的系统，其特征在于，所述转换过程还使所述处理器，在创建适配版本时，生成所述媒体数据的概要形式的所述源版本。

29. 如权利要求 21 所述的系统，其特征在于，所述多媒体表示的所述适配版本还包括对应于所述多媒体表示的复合数据（420）的源版本的适配的复合数据（440）。

30. 如权利要求 29 所述的系统，其特征在于，所述转换过程还使所述处理器在创建适配版本时，根据所述目标设备的能力和所述适配的媒体数据的特性，生成所述适配的复合数据。

31. 如权利要求 29 所述的系统，其特征在于，所述转换过程还使所述处理器在创建适配版本时，将所述多媒体表示的描述数据的源版本转换为所述描述数据的目标版本（212）；以及

根据所述描述数据的所述目标版本生成所述适配的复合数据（216）。

32. 如权利要求 29 所述的系统, 其特征在于, 所述转换过程还使所述处理器在创建适配版本时, 根据所述多媒体表示的描述数据的源版本导出所述复合数据的源版本 (208); 以及将所述复合数据的所述源版本转换为所述适配的复合数据 (210)。

33. 如权利要求 29 所述的系统, 其特征在于, 所述适配的复合数据包括所述适配的媒体数据中多个媒体对象的空间和时间布局以及同步信息。

34. 如权利要求 29 所述的系统, 其特征在于, 所述多媒体表示的所述源版本还包括所述复合数据的所述源版本。

35. 如权利要求 21 所述的系统, 其特征在于, 所述转换过程还使所述处理器在选择转换操作时, 将所选转换操作排序以满足最优化标准。

36. 如权利要求 21 所述的系统, 其特征在于, 所述转换操作是根据一组规则来选择的。

37. 如权利要求 21 所述的系统, 其特征在于, 所述目标设备的能力包括用于将所述适配的多媒体表示传送到所述目标设备的媒体的特性。

38. 如权利要求 21 所述的系统, 其特征在于, 所述转换过程还使所述处理器基于用户偏好来选择转换操作。

39. 如权利要求 21 所述的系统, 其特征在于还包括: 通过所述总线连接到所述处理器的接口 (736、720); 以及其中所述转换过程还使所述处理器将所述多媒体表示的所述适配版本通过所述接口传送到所述目标设备 (220)。

40. 如权利要求 21 所述的系统, 其特征在于还包括: 通过所述总线连接到所述处理器的接口 (736); 以及其中所述转换过程还使所述处理器通过所述接口接收所述多媒体表示的源版本的媒体数据源版本、复合数据源版本和描述数据源版本中的至少一项 (200, 202, 206)。

41. 一种用于转换多媒体数据以传送到多个目标设备的设备，包括：

用于根据目标设备（310）的能力从多个转换操作中选择转换操作的装置；以及

用以利用所选择的转换操作（320）并基于多媒体表示内的媒体数据的描述数据和复合数据，从多媒体表示的源版本创建适合于所述目标设备的所述多媒体表示的适配版本的装置，所述多媒体表示的适配版本包括对应于所述多媒体表示的媒体数据（302、304）的源版本的适配的媒体数据（322、320）。

42. 如权利要求 41 所述的设备，其特征在于，所述用于创建的装置包括：

用于将所述多媒体表示的描述数据的源版本转换为所述描述数据的目标版本（212）的装置；以及

用于根据所述描述数据的所述目标版本生成所述适配的媒体数据（216）的装置。

43. 如权利要求 41 所述的设备，其特征在于，所述用于创建的装置包括：

用于根据所述多媒体表示的描述数据的源版本导出所述媒体数据的源版本（208）的装置；以及

用于将所述媒体数据的所述源版本转换为所述适配的媒体数据（210）的装置。

44. 如权利要求 41 所述的设备，其特征在于，所述多媒体表示的所述适配版本还包括对应于所述多媒体表示的复合数据（420）的源版本的适配的复合数据（440）。

45. 如权利要求 44 所述的设备，其特征在于，所述用于创建的装置包括：

用于根据所述目标设备的能力和所述适配的媒体数据的特性生成所述适配的复合数据的装置。

46. 如权利要求 44 所述的设备，其特征在于，所述用于创建的装置包括：

用于将所述多媒体表示的描述数据的源版本转换为所述描述数据的目标版本（212）的装置；以及

用于根据所述描述数据的所述目标版本生成所述适配的复合数据（216）的装置。

47. 如权利要求 44 所述的设备，其特征在于，所述用于创建的装置包括：

用于根据所述多媒体表示的描述数据的源版本导出所述复合数据的源版本（208）的装置；以及

用于将所述复合数据的所述源版本转换为所述适配的复合数据（210）的装置。

48. 如权利要求 44 所述的设备，其特征在于，所述多媒体表示的所述源版本还包括所述复合数据的所述源版本。

49. 如权利要求 41 所述的设备，其特征在于还包括：用于将所述多媒体表示的所述适配版本传送到所述目标设备（220）的装置。

50. 如权利要求 41 所述的设备，其特征在于还包括：用于接收所述多媒体表示的所述源版本的媒体数据源版本、复合数据源版本和描述数据源版本中的至少一项（200、202、206）的装置。

对多媒体数据进行转换以将其传送到多个异种设备

相关申请

本申请要求于 2001 年 12 月 12 日提交的美国临时专利申请 No.60/340388 的利益，该申请通过引用结合到本文中。

技术领域

本发明涉及对多媒体数据的操作，更具体地说涉及转换多媒体数据以传送到多个异种目标设备。

版权注释/许可声明

本专利文档的部分公开内容含有受版权保护的资料。版权拥有者不反对按专利商标局专利文件或记录中的原样对本专利文档或专利公开内容制作传真形式的复制品，除此之外保留所有版权。如下注释适用于下文及附图中描述的软件和数据。Copyright ©2001，索尼电子公司，版权所有。

背景技术

随着数字设备(如个人计算机、数码相机、个人数字助理(PDA)、蜂窝电话、扫描仪等)的日益普及，按照众所周知的标准格式化的多媒体数据正为业余爱好者到新手到行家的所有社会成员所共享。规范多媒体数据的捕获、存储和传输的许多标准为数字设备厂商广泛接受，并逐渐与数字设备融合，以允许观看和共享多种格式和版本的多媒体数据。在因特网上，超文本标记语言(HTML)和同步多媒体集成语言(SMIL)是表示多媒体内容的通用标准。HTML 是万

维网联盟 (W3C) 定义的基于标准通用标记语言(SGML)的一种标准。HTML 将 Web 页面描述为一组多媒体对象、元素或资源, 如图像、视频、音频和 JAVA®应用程序连同表示结构。表示结构包括在因特网浏览器中显示 HTML 网页时预期的多媒体资源表示的有关信息。这包括例如不同多媒体元素的布局的有关信息。HTML 采用嵌套标记来表示表示结构。HTML 的更新的版本称为 XHTML, 它在功能上相当于基于 XML 而非 SGML 的 HTML 版本。SMIL 是基于 XML 的语言, 用于将不同的多媒体资源(如图像、视频、音频等)集成到单一表示内容中。SMIL 包含允许引用媒体资源并控制它们的包括定时和布局的表示内容的特征, 以及用于链接到其它表示内容以创建超媒体表示内容的特征。SMIL 是一种未针对表示中所用媒体资源或对象定义任何表示方式的复合语言。相反, SMIL 定义一组允许将多媒体资源或对象集成起来或组合到单个表示内容中的标记。虽然 HTML 中存在某些 SMIL 特征, 但 SMIL 侧重于媒体资源的空间和时间的布局, 并提供比 HTML 大的交互性控制。

表示多媒体内容的另一标准是由运动图像专家组第 4 版(这里称为 MPEG-4)定义的题为“音频可视对象的编码”的 ISO/IEC 14496 标准, MPEG-4 规定如何将可听、可视或视听内容表示为媒体对象, 其中每个媒体对象表示为单一基本流。在 MPEG-4 中, 媒体对象被组织起来形成视听场景。视听场景以结构化方式表示不同多媒体对象的复杂表示形式。在场景内, 多媒体对象可以是自然的(意味着是从自然界捕获的), 也可以是合成的(意味着是用计算机或其它装置生成的)。例如, 在 MPEG-4 中会通过文本、图像和音频流的媒体对象及描述如何组织这些对象的场景来描述具有音频背景的包含文本和图像的场景。MPEG-4 视听场景由组织成层次树结构(称为场景图形)的媒体对象组成。原始媒体对象(如静止图像、视频和音频)置于场景图形的叶部位。MPEG-4 对这些原始多媒体对象中许多对象, 如视频和音频的表示方式进行了标准化, 但并不限于配合

MPEG-4 规定的媒体表示方式来使用。每个媒体对象包含允许将该对象包含进视听场景中的信息。

在场景图形底部可找到作为树叶的原始媒体对象。更一般地说，MPEG-4 的场景描述可以在空间上将媒体对象置于二维（2D）和三维（3D）坐标系中，并应用转换（例如空间转换，旋转）来改变对象的表示方式，将原始媒体对象组成复合媒体对象，以及使场景内对象的表示同步。MPEG-4 场景描述是基于虚拟现实建模语言（VRML）的概念构建的。Web 3D 联盟为 VRML 场景定义了基于 XML 的表示方式，称为可扩展 3D（X3D）。虽然以优化二进制方式对 MPEG-4 场景进行了编码以便传输，但 MPEG 已为 MPEG-4 场景描述定义了基于 XML 的表示方式，称为可扩展 MPEG-4 文本格式（XMT）。XMT 利用基于 XML 的文本语法表示 MPEG-4 场景描述。

XMT 可以与 SMIL、VRML 和 MPEG-4 播放器互操作。XMT 格式可以直接由 SMIL 播放器解释并回放，并很容易地转换为 X3D 格式，之后由 X3D 或 VRML 播放器来回放。XMT 还可以编译为 MPEG-4 表示形式，如 MPEG-4 文件格式（称为 MP4），随后可以由 MPEG-4 播放器来播放。XMT 包含两种不同的格式：XMT-A 格式和 XMT-Ω 格式。XMT-A 是 MPEG-4 内容的基于 XML 的版本，它含有一个 X3D 子集，该子集带 XD3 扩展，以允许表示特定于 MPEG-4 的特征。XMT-A 在 MPEG-4 文本和二进制格式提供一一映射。XMT-Ω 是基于 SMIL 的 MPEG-4 场景的高级版本。

数字多媒体信息的发布和使用范围不断扩大，以致难以识别用户特别感兴趣的内容。各种组织尝试通过提供多媒体信息内容的描述来解决此问题。这种描述信息可用于搜索、过滤和/或浏览，以查找指定的内容。运动图像专家组（MPEG）颁布了通称为 MPEG-7 的多媒体内容描述接口标准，用于将多媒体信息的内容描述标准化。与在先的 MPEG 标准（包括定义如何表示编码多媒体内容的 MPEG-4）相反，MPEG-7 规定如何描述多媒体内容。

就内容的描述而言，MPEG-7 可用于描述 MPEG-4、SMIL、HTML、VRML 以及其它多媒体内容数据。MPEG-7 采用数据定义语言（DDL）（DDL 规定用于定义描述工具的标准集合和定义新描述工具的语言），并提供核心的描述符和描述模式集。描述符和描述模式集的 DDL 定义组织成用于不同内容类的“模式（schema）”。模式中的每个描述符的 DDL 定义规定对应特征的语法和语义。模式中每种描述模式的 DDL 定义规定其子组件、描述符和描述模式之间的关系的结构和语义。MPEG-7 DDL 的格式基于 XML 和 XML 模式标准，其中描述符、描述模式、语义、语法和结构用 XML 元素和 XML 属性来表示。

发明内容

转换多媒体表示以便在多个异种目标设备上播放。基于目标设备的能力来选择转换操作，并将选择的转换操作用于根据多媒体表示的源版本创建适配的多媒体表示版本。适配版本包含对应于多媒体表示的媒体数据的源版本的适配的媒体数据。在一个方面中，适配的多媒体表示版本还包含对应于多媒体表示的复合数据的源版本的适配的复合数据。在另一方面，适配的媒体数据是根据多媒体表示的描述数据的源版本创建的。

按照本发明的第一方面的一种用于转换多媒体数据以传送到多个目标设备的方法，包括：根据目标设备的能力从多个转换操作中选择转换操作；以及利用所述选择的转换操作并基于多媒体表示内的媒体数据的描述数据和复合数据，根据所述多媒体表示的源版本创建适合于所述目标设备的所述多媒体表示的适配版本，所述多媒体表示的所述适配版本包含对应于所述多媒体表示的媒体数据的源版本的适配的媒体数据。

按照本发明的第二方面的一种用于转换多媒体数据以传送到多个目标设备的系统，包括：通过总线连接到存储器的处理器；由所

述处理器从所述存储器执行的转换过程，所述转换过程使所述处理器根据目标设备的能力从多个转换操作中选择转换操作，利用所述选择的转换操作并基于多媒体表示内的媒体数据的描述数据和复合数据，根据多媒体表示的源版本创建适合于所述目标设备的所述多媒体表示的适配版本，所述多媒体表示的适配版本包含对应于所述多媒体表示的媒体数据的源版本的适配的媒体数据。

按照本发明的第三方面的一种用于转换多媒体数据以传送到多个目标设备的设备，包括：用于根据目标设备的能力从多个转换操作中选择转换操作的装置；以及用以利用所选择的转换操作并基于多媒体表示内的媒体数据的描述数据和复合数据，从多媒体表示的源版本创建适合于所述目标设备的所述多媒体表示的适配版本的装置，所述多媒体表示的适配版本包括对应于所述多媒体表示的媒体数据的源版本的适配的媒体数据。

附图说明

通过阅读下面详细说明并参考附图，将会清楚本发明的创新特征，附图中：

图 1 说明本说明书所述转换方法的概念示意图；

图 2A 说明本说明书所述转换方法的一个实施例所采用的操作流程图；

图 2B 说明本说明书所述转换方法的一个实施例所采用的操作流程图；

图 3 说明根据本说明书所述方法的适配过程的实施例的一个实例；

图 4 说明本说明书所述适配转换方法的一个特定实例；

图 5A 说明例示源多媒体表示数据；

图 5B 说明例示目标多媒体表示数据；

图 6A、6B 和 6C 说明例示转换规则；

图 7 说明可以实施本说明书所述转换和适配方法的实施例的环境。

具体实施方式

本说明书所述的转换允许对多媒体表示进行转换，以便传送到多个异种设备。多媒体表示可以包括媒体数据、复合数据和描述数据。在一个实施例中，本说明书所述的转换操作对源版本的媒体数据及（可选的）复合数据进行调整，以便可以在目标设备上或一类目标设备上播放所述多媒体表示。在另一个实施例中，源多媒体表示只包含可据以导出适配的媒体数据及（可选的）复合数据的描述数据。

为表示图像、音频和视频内容而定义的数据，如用于图像的众所周知的 GIF 和 JPEG 格式、用于音频的 MP3 和 WAV 格式以及用于视频的 MPEG-1 和 MPEG-2 格式及其它类似格式在本文中概括地称为媒体数据，而图像、视频或视频数据的单个实例称为媒体对象。其它标准规定有一些语言格式，用于定义如何在空间和时间上组合媒体对象以形成单一相关多媒体表示。这些标准，如运动图像专家组 MPEG-4（MPEG-4）标准、万维网联盟（W3C）同步媒体集成语言（SMIL）、虚拟现实建模语言（VRML）、可扩展 3D（X3D）、超文本标记语言（HTML）以及其它类似标准在本文中称为复合标准，结合这些标准的指令称为复合数据。复合数据规定媒体对象的空间和时间布局及其同步方式。复合数据连同复合数据引用的所有相关媒体数据这里被称为多媒体表示数据，多媒体表示数据的实例称为多媒体表示。可以独立于媒体数据的格式来选择复合数据的格式，因为复合数据格式是与媒体数据格式无关的。其它诸如 MPEG-7（全称为多媒体内容描述接口标准）的标准规定描述多媒体内容的格式。MPEG-7 标准所涵盖的数据通常称为元数据，即描述其它数据的数据。称为元数据且由 MPEG-7 及其它标准定义的数据这里称为描述

数据。描述数据可以在多媒体表示中与媒体数据和复合数据组合。在各种实施例中，包含多媒体表示数据的媒体数据、复合数据和描述数据以及多媒体表示数据本身可以其它众所周知的格式来表示。

本说明书所述的转换和适配操作可以自动或半自动方式对源多媒体表示（包括媒体数据、复合数据以及描述数据中的一项或多项）进行调整和转换，以便传送到多个异种目标设备上显示。适配是通过运用转换过程来实现的，所述转换过程作用于媒体数据、复合数据和描述数据的结构化表示、如 XML。这种适配过程可以对结构化复合数据表示，如 MPEG-4、XMT、SMIL、HTML 和 VRML/X3D 执行。可以按照 MPEG-7 标准来表示描述数据。适配过程可以通过一组重写或转换规则来实现，这些规则规定应该如何转换多媒体表示的复合数据、媒体数据和描述数据，以便在目标设备上显示。这些规则可以利用源媒体数据、源复合数据和/或源描述数据以及用户偏好或设备能力信息来确定如何执行适配过程。

图 1 说明本说明书所述转换方法的概念示意图。在一个实施例中，多媒体表示 100 可包含媒体数据 102、复合数据 104 和描述数据 106。多媒体数据 100 由转换引擎 110 处理，它参照每种型号、类型或类的目标设备的转换规则根据目标设备的能力来对包括媒体数据、复合数据和描述数据的多媒体表示进行调整。在一个实施例中，针对特定设备进行调整的各种规则可以插件模块的形式结合到转换引擎中。可以将源多媒体表示的适配版本传送到各种目标设备。例如，第一版本 120A 可以传送到第一设备 130A，第二版本 120B 可以传送到第二设备 130B，依此类推，直到可以传送到设备 N 130N 的第 N 版本 120N。

图 2A 说明本说明书所述转换方法的一个实施例所采用的操作流程图。该操作流程对应于图 1 所示的转换引擎 110 执行的操作。应理解，在不背离本发明范围的前提下，可以将或多或少的处理过程结合于图 2A 所示的方法及本说明书中所述的其它方法和过程中，并

且本说明书所述和所示的功能块的安排方式并不隐含任何特定顺序。在一个实施例中，如方框 200 所示，接收包含媒体数据、复合数据和描述数据的多媒体表示。在另一个实施例中，如方框 202 所示，可以接收包含媒体数据和复合数据的多媒体表示。在此实施例中，如方框 204 所示，可以从媒体数据和复合数据导出描述数据。从媒体数据导出描述数据可以根据题为“内容数据和描述数据之间的自动译码 (Transcoding between Content Data and Description Data)” (“'891 申请”) 的美国专利申请 No.10/114891 所述的方法来实现。如方框 210 所示，可针对每种目标设备或每一通用类的目标设备将包含媒体数据、复合数据和描述数据的多媒体表示转换成多个版本。更具体地说，根据定义每种目标设备所需的调整的规则，基于多媒体数据所要传送到目标设备的特征和能力将多媒体表示转换成多个目标版本。这样，目标版本适合于目标设备的能力。转换还可以基于转换系统和/或目标设备的用户偏好并由该用户偏好控制。适配的多媒体表示的相应版本传送到目标设备，如方框 220 所示。此传送操作可以自动进行，例如通过目标设备的订阅操作进行，或者可以应目标设备的特定传送请求而进行。

图 2B 说明本说明书所述转换方法的一个实施例所采用的操作流程图。在此实施例中，转换处理过程接收多媒体表示的描述数据，如方框 206 所示。在一个实施例中，转换处理过程直接对源描述数据操作。在此实施例中，源描述数据用于导出源媒体数据和源复合数据，如方框 208 所示。此转换操作根据对源描述数据操作的一组规则来控制。此转换操作可以通过各种方法来实现，包括利用'891 申请中所述的方法。在此实施例中，从源描述数据导出的源媒体数据可以从一个或多个媒体源获得。这些媒体源可位于本地位置，也可能位于需要通过一个或多个网络（例如因特网）进行通信的远程位置。根据每种目标设备的规则将所得的多媒体表示转换成多种目标版本（如方框 210 所示），以创建目标多媒体表示。转换还可以基

于转换系统和/或目标设备的用户偏好并由该用户偏好控制。适配的多媒体表示的适当版本被传送到目标设备，如方框 220 所示。

在另一个实施例中，源描述可以根据对应于每种目标设备的规则转换成目标描述数据，如方框 212 所示。目标描述数据描述要针对目标设备进行调整的媒体数据。对应于目标设备的目标复合数据和目标媒体数据是从目标描述数据生成的，如方框 216 所示。此操作可以通过各种方法来实现，包括利用'891 申请中所述的方法。在此实施例中，可以从一个或多个媒体源获得根据目标描述数据生成的目标媒体数据。这些媒体源可以位于本地位置，也可以位于需要通过一个或多个网络（例如因特网）进行通信的远程位置。将适配的多媒体表示的适当版本传送到目标设备，如方框 220 所示。

在一个实施例中，将所接收到的包含源描述数据、复合数据和源媒体数据的源多媒体以及导出的源描述数据，导出的源媒体数据和导出的源复合数据表示为基于 XML 的表示，如 SMIL 或称为 XMT-Q 的可扩展 MPEG-4 文本格式，XMT-Q 是采用 XML 的 MPEG-4 表示形式，类似于 SMIL。所描述的转换方法还可以应用于以其它二进制形式存储的 MPEG-4 数据，方法是利用众所周知的方法（如 XMT 的 MPEG-4 参考软件中公开的那些方法）将其转换为基于 XML 的表示，类似于 XMT。复合数据和描述数据都可以表示为 XML 文档。因此，适配过程是一个 XML 文档转换为另一个 XML 文档的过程。因而，在一个实施例中，适配处理实现为一组转换规则，这些规则作用于表示源描述数据、媒体数据和复合数据的 XML 数据结构，例如对复合数据应用 SMIL/XMT 数据和对描述数据应用 MPEG-7。转换多媒体表示的规则可以可扩展样式表语言（XSL）和可扩展样式表语言转换（XSLT）的扩展形式来编写。即，一个或多个 XSLT 文件可以控制如何转换多媒体数据，以便传送到目的设备并予以表示。

在一个实施例中，转换过程包括将一组转换规则应用于多媒体表示的描述数据。这些转换规则可以视为重写规则。每个规则可以

规定一个条件和操作对。每个规则的条件部分定义何时应用该规则，它是参照描述数据的结构化表示和目标设备能力表示的一部分来定义的。规则的操作部分根据源描述数据构造目标描述数据的一部分。转换过程通过重复应用其条件匹配的规则来进行，直到不再有匹配演化的描述数据的所述规则或者满足停止条件为止。停止条件在目标描述数据满足可在目标设备上显示的多媒体表示的描述要求时出现。在各种实施例中，应用规则的过程可以是确定的或非确定的。

在某些实施例中，可以将成本与每个规则相关联，以便可以利用本技术领域人员熟知的搜索和优化技术，应用搜索算法来查找最佳或接近最佳的规则序列，这些规则序列以最小的成本实现源描述转换。与规则对应的成本可以表示目标数据在何种程度上满足对表示进行的调整所针对的目标设备的要求。

当描述数据以 XML 方式表示或可以映射到等效的基于 XML 的表示时，转换可以利用 XSLT 编写的规则来实现，以及可以利用本技术领域人员熟知的技术通过 XSLT 引擎来实现。一旦通过转换过程创建了目标描述数据，就可以将'891 申请中所述的方法应用于将描述数据转译成目标媒体数据和目标复合数据。

目标媒体数据是通过将源媒体数据映射成目标描述数据中所描述的目标媒体的应用媒体适配方法从源媒体数据生成的。例如，当目标描述中的图像尺寸指定不同的图像尺寸时，就对图像应用对应的尺寸缩放操作。

在另一个实施例中，转换过程利用描述数据所控制的规则同时转换媒体数据和复合数据。此过程中所用的描述数据可以由外部提供，也可以自动生成。在此实施例中，转换过程由两种协同操作以对多媒体表示进行调整的转换操作组成：转换媒体数据的媒体转换；以及转换复合数据结构的复合转换。这种转换过程应用一序列的媒体和/或复合转换。

媒体转换可包括利用熟知的信号处理算法实现的低级操作，所

述信号处理算法例如执行格式转换的操作，如将图像从 JPEG 转换成 GIF 格式；或者改变媒体的低级属性的操作，如改变音频数据的采样率和缩放图像。其它媒体转换操作可以将媒体从一种格式转换成另一种，例如将视频转换成一序列表示媒体概要、如关键帧的图像。转换过程不依赖于源数据创作或创建实现的细节，但是需要知道目标媒体的格式。在一个实施例中，以插件组件的形式来实现原子媒体转换，其中所述插件组件输出描述由该插件组件实现的转换的标准接口。

复合转换作用于复合数据的结构化数据表示。这样的表示在采用诸如 SMIL、XMT 等复合数据格式时可以是基于 XML 的。复合转换还可以通过将其它表示转换成等效的基于 XML 的格式来实现。针对转换描述数据所述的类似技术还可用于实现复合转换。

在本说明书所述的转换方法的一个实施例中，将一组规则用于确定和控制对媒体和复合数据进行联合适应性调整。在此实施例中，每个规则指定一个条件和操作对。每个规则的条件部分定义何时对复合/媒体数据应用该规则，它是参照这里所引用的复合数据和媒体数据的复合数据和相关描述数据的结构化表示的一部分来定义的。操作部分运用媒体和复合调整，生成目标多媒体表示所需的目标复合数据结构和媒体数据。转换过程包括重复应用条件匹配的规则，直到不再有这样的规则可应用或出现停止条件为止。停止条件出现在目标复合数据和媒体数据满足可在目标设备上显示的多媒体表示的要求时。应用规则的过程可以是确定或非确定的。

在某些实施例中，可以将成本与每个规则相关联，以便可以利用本技术领域人员熟知的搜索和优化技术，应用搜索算法来查找最佳或接近最佳的规则序列，这些规则序列以最小的成本实现源描述转换。这种成本可以反映所得的输出目标数据在何种程度上满足对表示进行的调整所针对的目标设备的要求。

当复合数据以 XML 方式表示，或可以映射到等效的基于 XML

的表示时，转换可以利用 XSLT 编写的规则来实现以及利用本技术领域人员熟知的技术通过 XSLT 引擎来实现。

图 3 说明本说明书所述方法的适配过程的实施例的一个实例。多媒体表示 300 可包含音频数据 302 和视频数据 304 形式的媒体数据，音频数据 302 和视频数据 304 按照 MPEG-4/SMIL 树结构化格式中的复合数据进行编排。在一个实施例中，音频数据可以是 MP3 或其它熟知的音频格式，视频数据可以是 MPEG-4 视频或其它熟知的视频内容数据格式。除了媒体数据外，多媒体表示还可包含描述数据。转换引擎 310 接收多媒体数据并对其进行调整，以便可以传送并在各种目标播放设备 340 上播放或以其它方式显示。转换引擎 310 所执行的适配可以包括媒体转换，如当播放设备不能播放视频数据时将视频数据转换成一序列静止帧，如元素 324 所示。适配操作还可包括将语音转换成文本，如元素 322 所示。这样，调整的媒体数据可以适当的方式显示于目标设备上，并执行复合转换，如图元素 330 所示。即，可以将熟知格式（如 SMIL 或 HTML 等）的复合数据连同调整的媒体数据提供给目的设备，以便以符合特定调整处理的合理方式显示调整的媒体数据。例如，当把组合音视频片段中的多媒体内容调整为一序列静止帧和文本时，静止帧的显示必须与文本协调，以便观看者能够欣赏相应的显示。播放设备 340 可以包括电视机 342、PDA 344 和蜂窝电话 346。在一个实施例中，电视机可以接收多媒体数据的适配版本，该版本符合国家电视标准委员会（NTSC）、逐行倒相制式（PAL）、数字电视（DTV）以及其它类似标准，而提供给 PDA 和蜂窝电话的版本可以是源多媒体数据的降低了图像帧分辨率、降低了帧速率、减少了颜色数量等的降等级版本。

此外，降等级版本还可加以调整以减小多媒体数据的大小，从而适应借以将多媒体数据的适配版本传送到或以其它方式传递到目标设备的介质的带宽限制。例如，要通过蜂窝电话系统传送的数据

必须小于可以通过蓝牙或 IEEE 802.11 无线系统传送的数据，因为蜂窝电话系统的带宽较小。同样地，对应于符合 IEEE 802.11、802.11a、802.11b 和 802.11g 标准的各类目标设备，可以创建不同的适配版本。这样，适配的多媒体表示的保真度或质量可加以定制，以便与目标设备的通信流的能力和特性以及目标设备的分辨率、色彩和其它特征和功能相匹配。

图 4 说明本说明书所述适配转换方法的一个特定实例。在此实例中，源多媒体表示 410 可以是例如电视机 400 上显示的有关足球比赛的音视频馈送信号。此多媒体表示可以包括媒体数据、描述数据和复合数据。源复合数据 420 可以根据复合适配方法 426 来调整，以创建或导出适配的复合数据 440，视频数据 422 形式的媒体数据可以通过视频适配方法 424 来调整。更具体来说，如果要调整视频数据以便在 PDA 上显示，则可将每秒 40 帧、每帧 1200×1600 点/每英寸的源视频数据调整为或低等级为每秒 15 帧、每帧 20×30 点/每英寸，如降等级视频数据 428 所示。如果要适配局限性更大的目标设备（如蜂窝电话），则可以将视频数据调整成一序列静止帧，这些静止帧不同时刻的足球比赛画面。同样地，如果有与多媒体源表示内容相关联的音轨或语音信道，可以将该语音调整成文本。在此情况中，复合适配必须考虑文本与静止图像的协调和对齐，以在蜂窝电话上得到可以理解的表示形式。最终结果为目标 PDA 460 上显示的适配的目标多媒体表示 450。本段落所述的适配方式称为模态适配或模态转换。模态转换包括将媒体数据从源模态转换成目标模态，如从视频转换为静止图形，从第一语言转换为第二语言以及从语音转换为文本。

图 5A 说明例示源多媒体表示数据，而图 5B 说明例示目标多媒体表示数据。图 5A 和 5B 中的例示多媒体表示数据显示 SMIL 格式的复合数据。在这些示例中，出于说明的目的，简化了复合数据。所示源多媒体表示用于高性能设备，如个人计算机，所用语言为英

语。目标多媒体表示是调整源多媒体表示以适应功能低的设备（如 PDA）并将语言从英语改为日语的结果。更具体地说，图 5A 显示的是可以显示高质量的视频和音频的高性能设备的 SMIL 复合数据的摘录。该摘录是类似于图 4 所示的足球比赛的多媒体概要信息的一部分。图 5B 显示的是适用于性能较低的设备同样的摘录，该设备不能显示视频，且只可播放低质量音频。

图 5A 所示的源复合数据具有同时显示的三个媒体对象，如<par>元素 526 所示，它指示并行显示。第一个媒体对象（由<video>标记 520 指示）是 MPEG-2 视频，它来源于“r1”区所示数据源文件“soccer-goal-30fps.mpg”，每秒 30 帧且分辨率为 640x480 像素。第二媒体对象（由<audio>标记 522 指示）是 44 千赫兹的高质量英语 MP3 音频剪辑，它来源于源文件“narration-en-44khz.mp3”。第三个媒体对象 524 是英文源文件“caption-en.txt”的英语文字标题。

为了调整该源多媒体表示，源复合数据和源媒体数据都要经过转换，以得到图 5B 所示的目标多媒体表示。因为性能较低的目标设备不支持视频回放，所以所执行的第一适配操作将源视频数据转换成一组关键帧，选择这些关键帧用于概括视频内容。多媒体表示的这一部分用图 5B 中所示的“seq”和“img”标记 530 和 532 表示于复合数据中。在此实例中，还对音频进行调整，以便音频信号和音频内容都经过调整。因为较低质量的设备只支持低保真度音频回放，所以将源音频的格式从 MP3 调整为 WAV，并将采样率从 44 千赫兹降到 8 千赫兹，如 WAV 音频对象 534 所示。此外，音频对象和文本对象的语言从源语言英语调整为目标语言日语，如文本对象 536 所示。

图 6A、6B 和 6C 说明例示转换规则。这些规则提供了可用于实现从图 5A 所示的源多媒体表示数据到图 5B 所示的目标多媒体表示数据的转换规则的实例。图 6A、6B 和 6C 所示的规则以类似于 XSLT 的语言来表示。每个规则称为 XSLT 模板，表示转换（即重写）规

则，由<xsl:template>..
</xsl:template>语法指示，如 610A 和 610B 所示。“match”属性 612 所指示的规则的条件部分指明该规则适用的表示数据类型或种类。包括在“xsl:template”标记（如规则 R1 610 的标记 610A 和 610B）内的每个规则主体包含形成转换匹配规则条件的 SMIL 多媒体的一部分的结果的指令。

在图 6A 和 6B 中，规则 R1 至 R3 用于转换复合数据，称为复合数据转换规则，而在图 6B 中，规则 R4 至 R7 用于转换媒体数据，称为媒体数据转换规则。例示规则 R1 610 通过调用 VideoToKeyFrame（视频到关键帧转换）媒体转换规则，即图 6C 所示的规则 R4 680 来调整视频对象的组成以适应目标设备的性能。虽然未显示实现 VideoToKeyFrame 媒体转换的细节，但此转换规则通过从视频中选择一组关键帧创建了一序列概括该视频的图像。规则 R1 与图 5A 中所含的<video>元素 520 匹配，并将其转换成图 5B 中的<seq>。
</seq>数据 530。

例示规则 R2 620 通过根据与音频对象的媒体源相关联的描述数据应用转换操作，以调整源 SMIL 复合数据中的音频对象的组成。第一个条件 622 检查音频数据的采样率是否超过目标设备可支持的最大采样率 8 千赫兹。如果音频数据的采样率超过此值，则调用 AudioDownSample（音频向下采样）转换规则（如图 6C 的规则 R5 682），以便通过降低音频媒体数据的采样率来对该音频数据进行转换。例示规则 R2 按如下条件检查段 624 中指示采样率的描述数据：

“description(@src)//AudioCoding/Sample/@rate > 8000”。

段 624 所示的条件中所用的 description()函数返回由统一资源定位符(URL)指定的媒体对象的相关 MPEG-7 描述数据。第二个条件 626 中的类似测试检查音频数据是否是 WAV 格式的，如果不是 WAV 格式，则调用 AudioConvertFormat（音频转换格式）规则，如图 6C 的规则 R6 684 来进行格式转码。否则，该音频表示数据不经转码就直接通过。例示规则 R2 将应用于图 5A 所示的<audio>元素 522，以将

其转换为<audio>元素 534，其中的媒体数据（由“src”字段值变化指示）从 44 千赫兹的 MP3 格式转换为 8 千赫兹的 WAV 格式。

例示规则 R3 用于转换 SMIL 复合数据中的文本媒体对象的组成。例示规则 R3 630 包括条件 632，它检查文本语言是否是由 \$targetLanguage 变量指示的期望语言（假定从某些源已知），是否匹配该文本的语言。如果源语言与目标语言不匹配，则调用 TranslateText（翻译文本）转换规则（如图 6C 的规则 R7 686）将该文本转换为期望的目标语言。该规则可以应用于图 5A 所示的<text>元素 524，以翻译为图 5B 的<text>元素 536 所示的语言。

图 7 说明可以实施本说明书所述转换和适配方法的实施例的环境。这里公开的方法可以实现为软件、硬件以及软硬件组合（如固件）。媒体数据可以由一个或多个多媒体源生成、创作或以其它方式提供，以供服务器计算机 720 之用。在各种实施例中，媒体源可以是一个或多个数字电视广播、现场视频馈送、股票行情显示器、音频广播以及通过无线电波传送或在广域网（如因特网或其它类似网络 750）上广播的类似媒体源。在一个实施例中，本说明书所述的方法可以在计算机（如服务器计算机 720）上实现。在一个实施例中，服务器计算机 720 包括处理器 722 和存储器 724。在一个实施例中，实施本说明书所述方法的各种实施例的软件可以由处理器 722 来执行。处理器 722 可以是任何计算机处理器或微处理器，例如可从 Santa Clara（圣克拉拉州）的 Intel 公司得到的 Intel® Pentium® 4 处理器，并且存储器 724 可以为任何随机存取存储器（RAM）。网络接口 736 可以是模拟调制解调器、电缆调制解调器、数字调制解调器、网络接口卡和其它网络接口控制器，所述其它网络接口控制器允许通过广域网（WAN）如网络 750（诸如因特网）、通过局域网（LAN）、通过熟知的无线标准等进行通信。

在一个实施例中，可以将软件程序形式的计算机指令存储在存储设备 726 上，该存储设备可以是硬盘驱动器。在一个实施例中，

可以实现本说明书所述方法的软件可以称为转换软件 728。此计算机软件可以通过网络 750 或其它通过网络接口 736 的 WAN 或 LAN 下载到服务器计算机 720，并存储在存储器 724 和/或存储设备 726 中。在各种实施例中，存储设备 726 可以是任何机器可读媒体，包括磁存储设备如硬盘驱动器、软盘驱动器以及光存储设备（如只读光盘存储器（CD-ROM）和可读可写光盘（CD-RW）设备）、可读可写多功能数字盘（DVD）设备、RAM、只读存储器（ROM）、闪存设备、存储棒设备、可以电方式擦写的可编程只读存储器（EEPROM）以及其它硅设备。在各种实施例中，一个或多个机器可读媒体可以是在本地连接的，如存储设备 726，也可以是通过电气方式、光学方式、无线方式、声学方式以及其它方式（包括通过网络）从远程源来访问的。

在一个实施例中，处理器 722、存储器 724、存储设备 726、USB 控制器 730 和网络接口 736 分别连接到总线 740，这些设备通过总线 740 可彼此进行通信。在各种实施例中，服务器计算机中可以包括两条或更多的总线。此外，在各种实施例中，服务器计算机 720 的两个或两个以上的组件可以包括在服务器计算机 720 中。众所周知，服务器计算机 720 可包括可从 Microsoft 公司（Redmond, Washington）获得的 Microsoft® Windows® XP 专业版的操作系统。

在一个实施例中，服务器计算机 720 可以实现为两个或两个以上的计算机，这些计算机配置成一个群集、一组、局域网（LAN）、子网或其它的多计算机组织形式。此外，当由多个计算机组成时，服务器计算机组可以包括路由器、集线器、防火墙以及其它联网设备。在此实施例中，所述组可包括多个专用服务器，如图形服务器、音频服务器、事务服务器、应用服务器等。在一个实施例中，服务器计算机 720 可以依靠一个或一个以上的第三方（未显示），通过网络 750 或直接连接来提供事务处理和/或其它信息以及处理帮助。

在一个实施例中，作为目的地设备 760 显示的目标计算设备，

如个人计算机、个人数字助理 (PDA)、蜂窝电话、计算输入板 (computing tablet)、便携式计算机等的用户可以通过网络 750 与服务器计算机 720 进行通信, 以获得来自远程源如多媒体源 710 的多媒体数据。在一个实施例中, 目的地设备 760 可以具有类似于服务器计算机 720 的配置。此外, 目标设备包括视频显示装置和/或音频输出装置, 这些装置在各种实施例中允许目标设备的用户查看诸如视频、图形和/或文本的信息, 以及听取各种质量的音频, 所有这些都取决于目标设备的视频显示装置和音频装置的性能。目标设备还包括用户输入装置, 如键盘、小键盘、触摸屏、鼠标、输入笔等。

在一个实施例中, 在根据本说明书所述方法对多媒体表示的复合数据、描述数据和/或媒体数据进行转换和调整之后, 服务器计算机 720 可以获得多媒体表示数据并将其传送到本地设备 770。本地设备可以是蜂窝电话、PDA、MP3 播放器、便携式视频播放器、便携式计算机以及可以通过如下方式接收转换的多媒体表示数据和媒体数据的类似设备: 电气方式、光学方式、无线方式、声学方式以及符合任何熟知的通信标准的其它方式, 通信标准包括例如通过 USB 控制器 730 的通用串行总线(USB)、IEEE 1394(更一般地称为 I.Link® 和 Firewire®)、Bluetooth™等。服务器 720 和本地设备之间的通信可以支持诸如如下的通信协议: HTML、IEEE 802.11、用于移动设备的 W3PP 和/或 WAP 协议以及用于请求多媒体表示数据的其它熟知的通信协议。

以上说明参照本发明的具体实施例来进行描述。显然, 在不背离所附权利要求书阐明的本发明的精神和范围的前提下, 可以对那些实施例进行各种修改和变更。因此, 本说明书和附图应视为说明性而非限定性的。

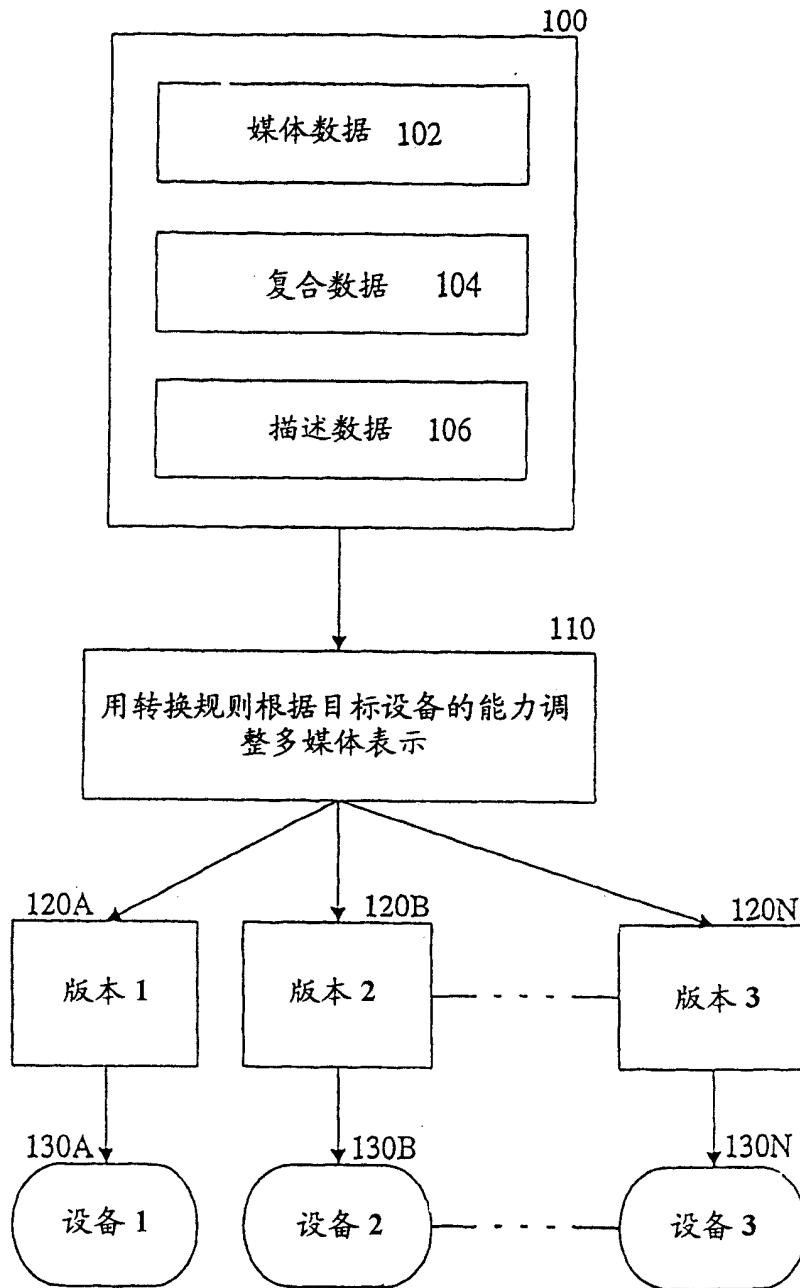


图 1

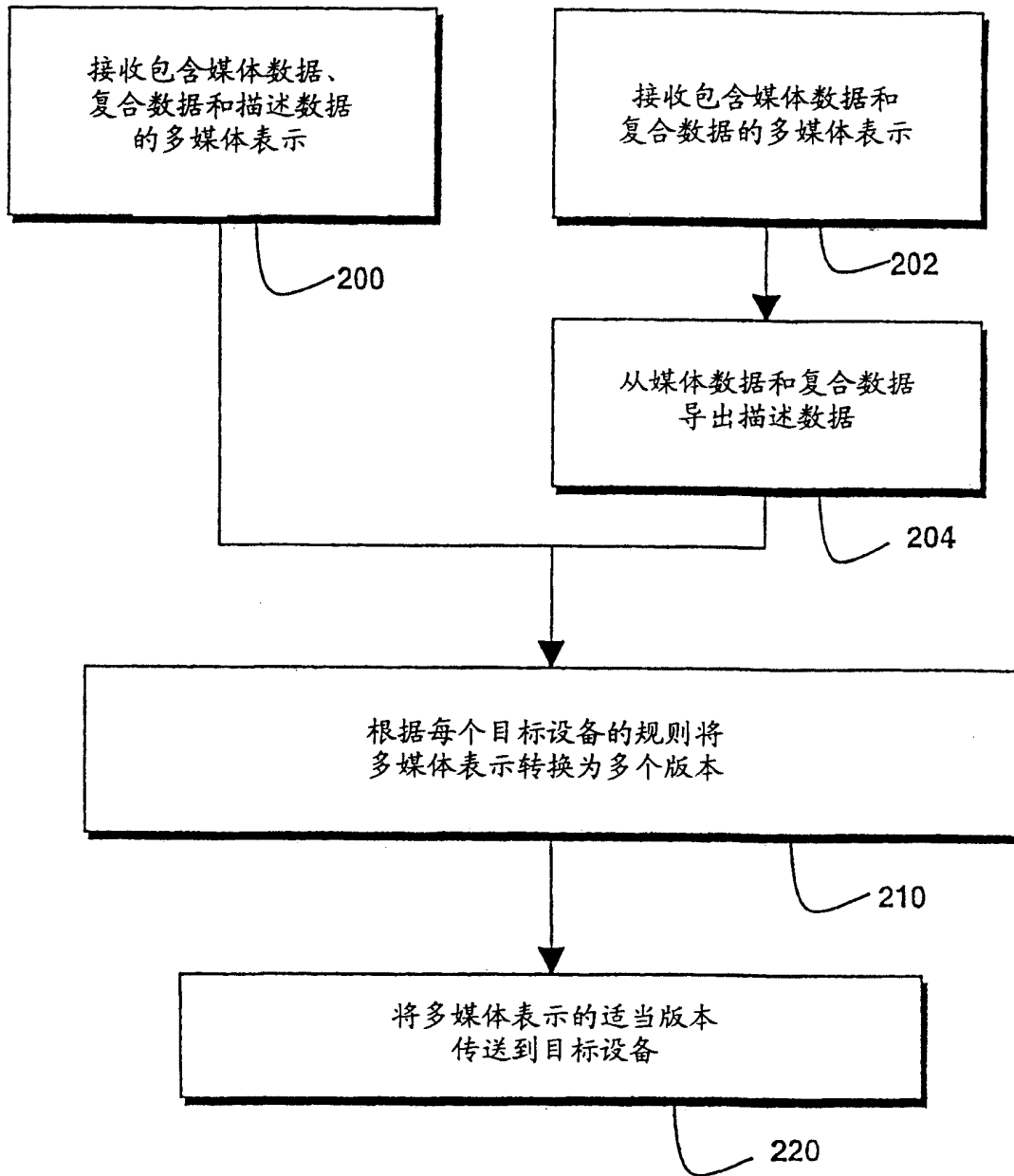


图 2A

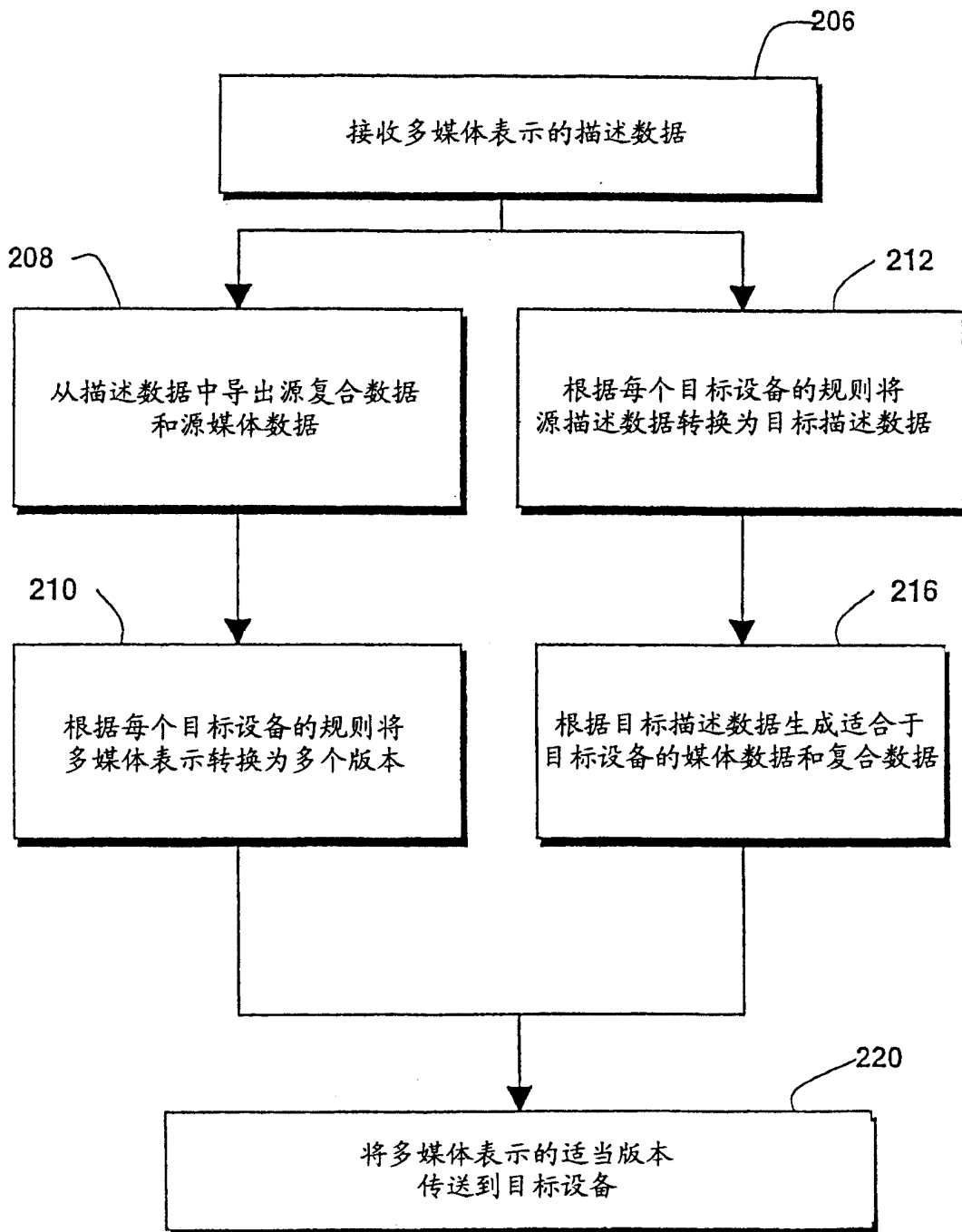


图 2B

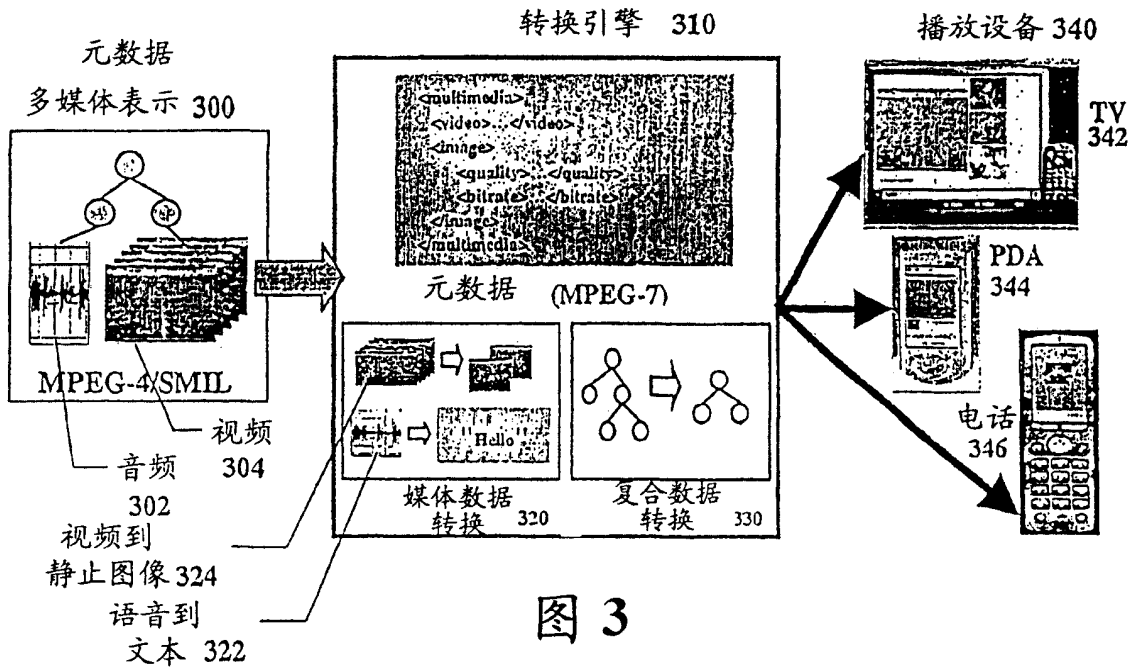


图 3

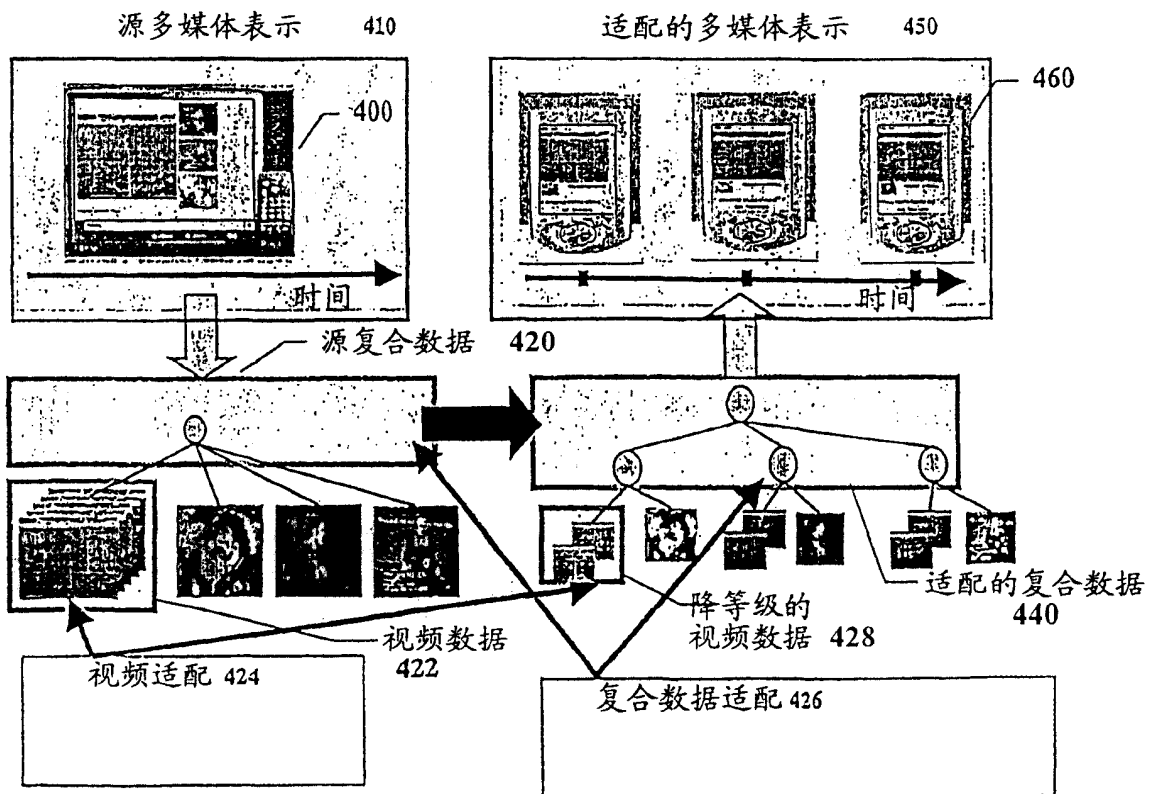


图 4

```

<smil>
  <head>
    <!-- 细节省略 -->
    <region id="r1" left="0" top="0" right="640"
bottom="480"/>
    <region id="r2" left="100" top="480" right="300"
bottom="520"/>
  </head>
  <body>
    <!-- 场景数据 -->
    <par>
      <video region="r1" src="soccer-goal30fps.mpg"/> 520
      <audio src="narration-en-44khz.mp3"/> 522
      <text region="r2" src="caption-en.txt"/> 524
    </par>
    <!-- 更多场景数据 -->
  </body>
</smil>

```

图 5A

```

<smil>
  <head>
    <!-- 细节省略 -->
    <region id="r1" left="0" top="0" right="320"
bottom="240"/>
    <region id="r2" left="50" top="320" right="150"
bottom="340"/>
  </head>
  <body>
    <!-- 场景数据 -->
    <par>
      <seq> 532
        
        
        
        <!-- additional image may be listed here -->
      </seq> 534
      <audio src="narration-ja-8khz.wav"/> 536
      <text region="r2" src="caption-ja.txt"/>
    </par>
    <!-- 更多场景数据 -->
  </body>
</smil>

```

图 5B

```

<!------- 复合数据转换 ----->
<!-- 规则 R1 -->
<xsl:template match="video">
  <!-- 将视频转换为序列关键帧 -->
  <seq>
    <xsl:call-template name="VideoToKeyFrame">
      <xsl:with-param name="video" select="."/>
    <xsl:call>
  </seq>
</xsl:template>
<!-- 规则 R2 -->
<xsl:template match="audio">
  <xsl:choose>
    <!--如果采样率大于8千赫兹则进行调整-->
    <xsl:when test="description(@src)//AudioCoding/Sample/@rate
      > 8000">
      <!-- 降低媒体的采样率到8千赫兹 -->
      <xsl:call-template name="AudioDownSample">
        <xsl:with-param name="source" select="@src"/>
        <xsl:with-param name="rate" select="8000"/>
      </xsl:call-template>
    </xsl:when>
    <xsl:when test="description(@src)//AudioCoding/Format !=
      'WAV' ">
      <xsl:call-template name="AudioConvertFormat">
        <xsl:with-param name="source" select="@src"/>
        <xsl:with-param name="format" select="wav"/>
      </xsl:call-template>
    </xsl:when>
    <!-- 其他条件和转换 -->
  </xsl:choose>
</xsl:template>

```

图 6A


```

    630
<!-- 规则 R3 -->
<xsl:template match="text">
  <xsl:choose>
    <xsl:when select="description(@src)//Language !=
    $targetLanguage"> 632
      <xsl:call-template name="TranslateText">
        <xsl:with-param name="source" select="@src"/>
        <xsl:with-param name="to" select="$targetLanguage"/>
      </xsl:call-template>
    </xsl:when>
    <xsl:otherwise>
      <xsl:copy-of select="."/>
    </xsl:otherwise>
  </xsl:choose>
</xsl:template>

```

图 6B

```

<!------- 媒体数据转换 ----->
    680
<!-- 规则 R4 -->
<xsl:template name="VideoToKeyFrame">
  <!-- 细节省略 -->
</xsl:template>
    682
<!-- 规则 R5 -->
<xsl:template name="AudioDownSample">
  <!-- 细节省略 -->
</xsl:template>
    684
<!-- 规则 R6 -->
<xsl:template name="AudioConvertFormat">
  <!-- 细节省略 -->
</xsl:template>
    686
<!-- 规则 R7 -->
<xsl:template name="TranslateText">
  <!-- 细节省略 -->
</xsl:template>

```

图 6C

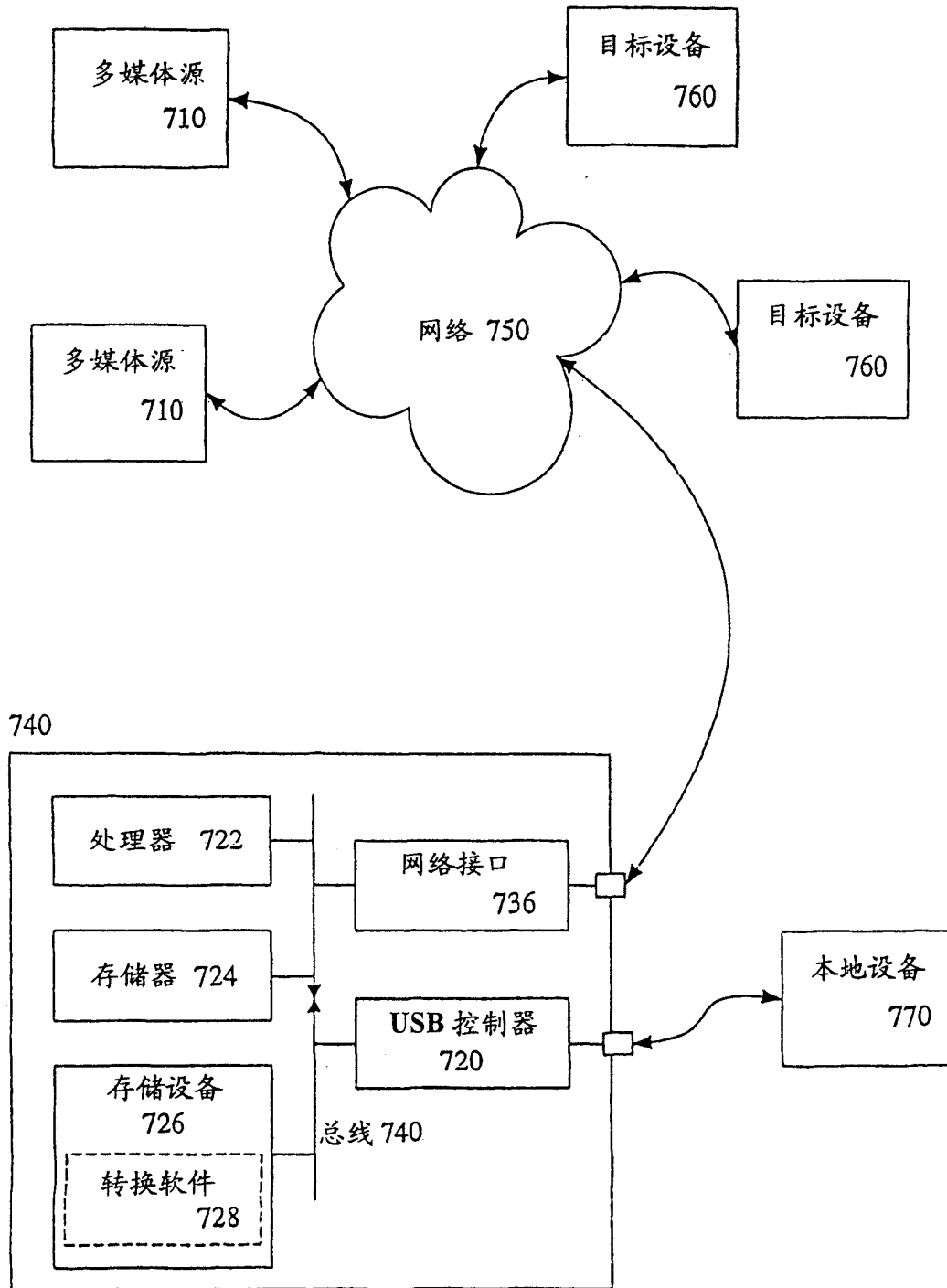


图 7