[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
H04N 5/44 (2006.01)

[12] 发明专利说明书

专利号 ZL 03811689.8

[45] 授权公告日 2007 年 11 月 28 日

[11] 授权公告号 CN 100352266C

[22] 申请日 2003.5.19 [21] 申请号 03811689.8

[30] 优先权

[32] 2002.5.20 [33] US [31] 10/151,599

[86] 国际申请 PCT/US2003/015802 2003.5.19

[87] 国际公布 WO2003/101093 英 2003.12.4

[85] 进入国家阶段日期 2004.11.22

[73] 专利权人 汤姆森特许公司

地址 法国布洛涅

[72] 发明人 约翰·P·曼宁

[56] 参考文献

JP2001-94857A 2001.4.6

WO0219309A1 2002.3.7

US-5045939A 1991.9.3

US5500683A 1996.3.19

US-5500683A 1996.3.19

US-5999220A 1999.12.7

审查员 吴黄飞

[74] 专利代理机构 北京市柳沈律师事务所

代理人 吕晓章 马 莹

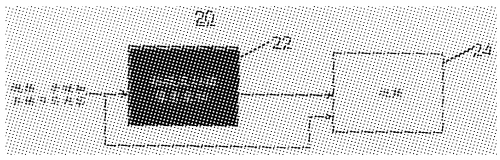
权利要求书 3 页 说明书 7 页 附图 1 页

[54] 发明名称

视频图像格式化技术

[57] 摘要

一种建立摇摄及扫描像素坐标(P)的方法，该坐标定义用于格式化具有第一宽高比的图像以产生具有第二宽高比的图像的图像捕获视窗，该方法包括如下步骤：通过检验与图像帧相关联的视频、音频和图像数据来检验输入的图像帧以确定其中最相关动作的位置；和根据以下关系式，按照视频、音频和图像数据的预定组合，根据最相关动作在输入图像帧中的位置来建立摇摄及扫描像素坐标：P = X1 * F(V) + X2 * F(A) + X3 * F(D)。通过在定义用于格式化的图像捕获视窗的图像中建立摇摄及扫描像素坐标(P)，对具有第一宽高比的视频图像帧(10)格式化，从而生成具有不同宽高比的图像帧。预测器(22)通过检验图像帧来确定图像中最相关动作的位置以确定摇摄及扫描像素坐标。预测器(22)通过检验视频数据、音频数据以及诸如封闭标题信息的其它数据之一来确定最相关动作的位置。



1．一种建立摇摄及扫描像素坐标(P)的方法，该坐标定义用于格式化具有第一宽高比的图像以产生具有第二宽高比的图像的图像捕获视窗，该方法包括如下步骤:

通过检验与图像帧相关联的视频、音频和图像数据来检验输入的图像帧以确定其中最相关动作的位置；和

根据以下关系式，按照视频、音频和图像数据的预定组合，根据最相关动作在输入图像帧中的位置来建立摇摄及扫描像素坐标:

P=X1*F(V)+X2*F(A)+X3*F(D)

其中，X1=分配给视频数据的加权因子(0<X1<1);

X2=分配给音频数据的加权因子(0<X2<1);

X3=分配给图像数据的加权因子(0<X3<1);

F(V)=使用视频数据预测的坐标;

F(A)=使用音频数据预测的坐标;

F(D)=使用图像数据预测的坐标;

并且，X1+X2+X3=1。

2．一种建立摇摄及扫描像素坐标(P)的装置，该坐标定义用于将图像从第一宽高比格式化成第二宽高比的图像捕获视窗，该装置包括:

用于通过检验与图像帧相关联的视频、音频和图像数据来检验图像帧以确定其中最相关动作的位置的部件;

用于按照视频、音频和图像数据的预定组合根据最相关动作在图像中的位置来建立摇摄及扫描像素坐标的部件，

其中所述确定部件根据以下关系式建立摇摄及扫描像素坐标(P):

P=X1*F(V)+X2*F(A)+X3*F(D)

其中，X1=分配给视频数据的加权因子(0<X1<1);

X2=分配给音频数据的加权因子(0<X2<1);

X3=分配给图像数据的加权因子(0<X3<1);

F(V)=使用视频数据预测的坐标;

F(A)=使用音频数据预测的坐标;

F(D)=使用图像数据预测的坐标;

并且，X1+X2+X3=1。

3. 一种对具有第一宽高比的输入图像帧格式化从而生成具有第二宽高比的格式化的图像帧的方法，包括如下步骤：

通过检验与图像帧相关联的视频、音频和图像数据来在图像帧中建立摇摄及扫描像素坐标(P)，该坐标基本与图像帧中与最相关动作相关联的位置重合，和

按照视频、音频和图像数据的预定组合，用以摇摄及扫描像素位置为中心的图像捕获视窗来摇摄及扫描输入图像帧，从而格式化输入图像帧以产生格式化的图像数据帧，

其中根据以下关系式确定所述摇摄及扫描像素坐标(P)：

P=X1*F(V)+X2*F(A)+X3*F(D)

其中，X1=分配给视频数据的加权因子(0<X1<1);

X2=分配给音频数据的加权因子(0<X2<1);

X3=分配给图像数据的加权因子(0<X3<1);

F(V)=使用视频数据预测的坐标;

F(A)=使用音频数据预测的坐标;

F(D)=使用图像数据预测的坐标;

并且，X1+X2+X3=1。

4. 一种对具有第一宽高比的输入图像帧格式化从而生成具有第二宽高比的格式化的图像帧的装置，该装置包括：

用于通过检验与图像帧相关联的视频、音频和图像数据来在图像帧中建立摇摄及扫描像素坐标(P) 的部件，该坐标基本和图像帧中的、与最相关动作相关联的位置重合，和

用于按照视频、音频和图像数据的预定组合，用以摇摄及扫描像素坐标为中心的图像捕获视窗来摇摄及扫描输入图像帧，从而格式化输入图像帧以产生格式化的图像帧的部件，

其中所述建立部件根据以下关系式建立摇摄及扫描像素坐标(P)：

P=X1*F(V)+X2*F(A)+X3*F(D)

其中，X1=分配给视频数据的加权因子(0<X1<1);

X2=分配给音频数据的加权因子(0<X2<1);

X3=分配给图像数据的加权因子(0<X3<1);

F(V)=使用视频数据预测的坐标;

F(A)=使用音频数据预测的坐标;

F(D)=使用图像数据预测的坐标;

并且，X1+X2+X3=1。

视频图像格式化技术

本申请是非临时申请，要求根据 35 U.S.C.§119 的、于 2002 年 5 月 20 日申请的题目为“VIDEO IMAGE FORMATTING TECHNIQUE”的第 10/151599 号美国国家非临时申请的权益，这里引用其内容作为参考。

技术领域

本发明涉及一种格式化视频图像尺寸的技术，以便以不同尺寸在显示设备上处理或显示该视频图像。

背景技术

以前，美国的国家电视系统委员会(NTSC)对电视图像采用 4:3 的宽高比，因为早期的电影工业在判定电影观众更倾向于宽比高更大的矩形图像后，对电影采用了同样的宽高比。全世界其它国家的电视标准机构也采用了 4:3 的宽高比。随着高清电视的出现，例如由 Grand Alliance 提议且在 1994 年 3 月 20 日至 24 日举行的全国广播工作者联合会、第 48 届年度广播工程会议的会议纪要中被描述的高清陆地播放系统，内容供应商现在提供在宽屏电视显示设备上放映宽高比为 16:9 的高清电视节目。

虽然市场对高清宽屏电视显示设备的需求持续增长，但仍有大量潜在基数的标准电视显示设备只能显示 4:3 宽高比的图像。为了便于格式化具有 16:9 宽高比的宽屏视频图像以便在仅能显示 4:3 宽高比的电视显示设备上显示，宽屏内容的供应商通常在每一单个帧之内编码像素坐标以便能由“摇摄及扫描(pan and scan)”操作进行图像格式转换。同样地，今后几年 16:9 宽高比显示设备将占据更大份额，但在 4:3 宽高比时代产生的视频内容仍将继续播放。因此，同样存在将 4:3 宽高比图像剪辑为适应于宽高比为 16:9 的视频的需求。

术语“摇摄及扫描”是指一种众所周知的技术，用于通过在视窗内有效地摇摄一行图像元素(像素)然后扫描下一行像素而开始随后的摇摄等操作，在数字宽屏视频图像帧内捕获固定尺寸视窗。如上所述，与摇摄及扫描格式操作相关的图像拍摄视窗具有指定尺寸。因此，单像素坐标，一般是指“摇摄

及扫描像素坐标”将定义这样一个视窗。实际上，摇摄及扫描像素坐标定义视窗的中心。摇摄及扫描像素坐标建立的位置影响格式化后宽屏图像所保留的部分。例如，在宽屏图像中心建立摇摄及扫描像素坐标，一般情况下所产生的格式化图像将不包含任何在图像边缘发生的动作。为了避免这样的难点，一些视频图像内容供应商嵌入摇摄及扫描坐标来以便于格式化。然而，由于视频图像是动态的，所以这些供应商们通常必须在每一个连续的图像帧内嵌入摇摄及扫描坐标以便能正确格式化。已经证实提供这样的摇摄及扫描坐标是相当繁重的过程，因而只有极少数供应商能真正地做到这一点。

因此，需要一种自动建立摇摄及扫描坐标的技术来克服在先技术的缺点。

发明内容

简单地说，根据本发明的原理，提供一种方法，用于在一个图像帧内建立像素坐标(即，“摇摄及扫描”像素坐标)，该坐标定义用于将图像帧从第一宽高比(即，第一尺寸)格式化成第二宽高比(即，第二尺寸)的图像捕获视窗。依照此方法，对图像帧进行检验以确定其中最相关动作的位置。现存几种可能的技术用于在图像帧内建立最相关动作的位置。例如，在 MPEG 格式图像被编码成视频图像帧的情况下，坐标与展示最大变化的 I，P，B 画面里的宏块相关的坐标，或展示最大变化的移动向量的位置，能在图像帧内定义最相关动作的坐标。另外，在其比特率展示最大变化率的像素的视频图像帧内的位置也能够构成最相关动作的位置。对应于声音的音频数据、音响效果或是与图像源相关的环绕音，也能建立最相关动作位置，而并不只依赖于视频数据。另外，或者除了使用视频数据和/或音频数据外，像图像中封闭字幕信息这样的其它数据也能在图像中建立最相关动作位置。实际上，图像中最相关动作位置可以由视频数据、音频数据及其它数据的加权和给出。在图像中建立最相关动作位置后，建立摇摄及扫描像素坐标以便其定位与图像中最相关动作的位置重合。

根据本发明另一个方面，提供一种技术，用于把具有第一宽高比的图像格式化为具有第二宽高比的图像。根据这种方法，根据最相关动作图像中的位置来建立摇摄及扫描像素坐标，即，使坐标定位。实践中，对于电视图像，一个或多个视频图像数据、音频数据和/或其它数据的组合被处理以在图像中产生最相关动作的位置。摇摄及扫描被定位成与图像中的最相关动作位置重

合。摇摄及扫描坐标定义图像捕获视窗的位置，当被与摇摄及扫描格式化操作相关联地应用时，其能方便地捕获具有所希望宽高比的图像。

按照本发明的一个方面，提供一种建立摇摄及扫描像素坐标(P)的方法，该坐标定义用于格式化具有第一宽高比的图像以产生具有第二宽高比的图像的图像捕获视窗，该方法包括如下步骤：通过检验与图像帧相关联的视频、音频和图像数据来检验输入的图像帧以确定其中最相关动作的位置；和根据以下关系式，按照视频、音频和图像数据的预定组合，根据最相关动作在输入图像帧中的位置来建立摇摄及扫描像素坐标：P=X1*F(V)+X2*F(A)+X3*F(D)，其中，X1=分配给视频数据的加权因子(0<X1<1)；X2=分配给音频数据的加权因子(0<X2<1)；X3=分配给图像数据的加权因子(0<X3<1)；F(V)=使用视频数据预测的坐标；F(A)=使用音频数据预测的坐标；F(D)=使用图像数据预测的坐标；并且，X1+X2+X3=1。

按照本发明的另一方面，提供一种建立摇摄及扫描像素坐标(P)的装置，该坐标定义用于将图像从第一宽高比格式化成第二宽高比的图像捕获视窗，该装置包括：用于通过检验与图像帧相关联的视频、音频和图像数据来检验图像帧以确定其中最相关动作的位置的部件；用于按照视频、音频和图像数据的预定组合根据最相关动作在图像中的位置来建立摇摄及扫描像素坐标的部件，其中所述确定部件根据以下关系式建立摇摄及扫描像素坐标(P)：P=X1*F(V)+X2*F(A)+X3*F(D)，其中，X1=分配给视频数据的加权因子(0<X1<1)；X2=分配给音频数据的加权因子(0<X2<1)；X3=分配给图像数据的加权因子(0<X3<1)；F(V)=使用视频数据预测的坐标；F(A)=使用音频数据预测的坐标；F(D)=使用图像数据预测的坐标；并且，X1+X2+X3=1。

按照本发明的另一方面，提供一种对具有第一宽高比的输入图像帧格式化从而生成具有第二宽高比的格式化的图像帧的方法，包括如下步骤：通过检验与图像帧相关联的视频、音频和图像数据来在图像帧中建立摇摄及扫描像素坐标(P)，该坐标基本与图像帧中与最相关动作相关联的位置重合，和按照视频、音频和图像数据的预定组合，用以摇摄及扫描像素位置为中心的图像捕获视窗来摇摄及扫描输入图像帧，从而格式化输入图像帧以产生格式化的图像数据帧，其中根据以下关系式确定所述摇摄及扫描像素坐标(P)：P=X1*F(V)+X2*F(A)+X3*F(D)，其中，X1=分配给视频数据的加权因子(0<X1<1)；X2=分配给音频数据的加权因子(0<X2<1)；X3=分配给图像数据的

加权因子(0<X3<1)；F(V)=使用视频数据预测的坐标；F(A)=使用音频数据预测的坐标；F(D)=使用图像数据预测的坐标；并且，X1+X2+X3=1。

按照本发明的另一方面，提供一种对具有第一宽高比的输入图像帧格式化从而生成具有第二宽高比的格式化的图像帧的装置，该装置包括：用于通过检验与图像帧相关联的视频、音频和图像数据来在图像帧中建立摇摄及扫描像素坐标(P) 的部件，该坐标基本和图像帧中的、与最相关动作相关联的位置重合，和用于按照视频、音频和图像数据的预定组合，用以摇摄及扫描像素坐标为中心的图像捕获视窗来摇摄及扫描输入图像帧，从而格式化输入图像帧以产生格式化的图像帧的部件，其中所述建立部件根据以下关系式建立摇摄及扫描像素坐标(P)：P=X1*F(V)+X2*F(A)+X3*F(D)，其中，X1=分配给视频数据的加权因子(0<X1<1)；X2=分配给音频数据的加权因子(0<X2<1)；X3=分配给图像数据的加权因子(0<X3<1)；F(V)=使用视频数据预测的坐标；F(A)=使用音频数据预测的坐标；F(D)=使用图像数据预测的坐标；并且，X1+X2+X3=1。

附图说明

图 1 描述了宽屏视频图像的帧；和

图 2 描述了根据本发明原理的用于建立摇摄及扫描像素坐标的位置以便使图 1 中的图像格式化的装置的示意性框图。

具体实施方式

图 1 描述了具有用于在宽屏电视显示设备(未示出)上显示的宽高比为16:9 的典型宽屏视频图像帧 10。重叠在图像帧 10 上的是第二图像帧 12，第二图像帧 12 的大小对应于“标准”图像帧，其具有用于在标准电视显示设备(未示出)上显示的 4:3 的宽高比。在未格式化时，宽屏图像帧 10 对于在标准电视显示设备上显示来说太大了。因而，宽屏图像帧 10 必须经格式化以形成标准大小的视频图像 12。

现存的各种技术中，用于把宽屏视频图像帧 10 格式化成标准大小(4:3 的宽高比)的视频图像帧 12 的是“摇摄及扫描”技术。摇摄及扫描技术能通过建立大小与标准图像帧 12 相同的图像捕获视窗来格式化宽屏图像帧(即，图像帧 10)。图像捕获视窗中的各行像素被连续摇摄，之后，进行扫描操作以跳

至下一行像素来进行后面的摇摄操作。因为与摇摄及扫描格式化操作相关的图像捕获视窗具有固定大小，所以单个像素坐标(以下称为“摇摄及扫描像素坐标”)可以定义该视窗。实践中，摇摄及扫描像素坐标定义图像捕获视窗的中心。

以前，有些宽屏视频图像内容供应商嵌入摇摄及扫描像素坐标以便通过摇摄及扫描技术进行格式化。由于宽屏视频图像 10 中的内容可能是动态的，所以摇摄及扫描像素坐标的位置必须改变以确保正确格式化以后得到的最终图像反映宽屏图像中的内容。在有些情况下，每一帧中摇摄及扫描坐标都要改变，由于这个原因，很多内容供应商不提供摇摄及扫描像素坐标，因为这样做需要时间和精力。

图 2 示出了根据本发明原理的装置 20，用于自动建立摇摄及扫描像素坐标以便通过摇摄及扫描格式化操作对类似图 1 中所示的宽屏图像 10 这样的图像进行格式化。装置 20 包括预测器 22，用于根据图像中最相关动作的位置来确定图像的摇摄及扫描像素坐标。预测器 22 一般包括处理器、数字信号处理器(DSP)、或固定逻辑阵列(例如可编程阵列逻辑(PAL)或现场可编程门阵列(FPGA))、或者这些元件的组合，用于对一个或多个视频数据、音频数据或其他数据(例如封闭标题数据)进行处理以从这些数据建立图像中的最相关动作的位置。根据图像内的最相关动作位置的信息可以建立摇摄及扫描像素坐标使其与该位置重合。

例如，在 MPEG 编码的视频数据的情况下，预测器 22 根据 I、P 和 B 画面中的宏块坐标建立摇摄及扫描像素坐标。另外，在 MPEG 编码的视频数据的情况下，预测器 22 可以根据运动矢量表示最大变化的位置建立摇摄及扫描像素坐标。根据其它可替换的方法，预测器 22 可以根据具有最高比特率的图像中的像素的位置建立摇摄及扫描像素坐标。

预测器 22 可以使用音频数据来确定摇摄及扫描像素坐标，而不是使用视频数据。例如，预测器 22 可以使用数字和/或模拟音频数据来确定图像中与声音(左或右)的位置或具有最大量级的音响效果的位置相应的最相关动作位置。预测器 22 还可以利用环绕声音(中间、左、右或环绕左/右)的位置来确定图像中最相关动作的位置并建立与该位置重合的摇摄及扫描像素坐标的位置。

预测器 22 还可以利用诸如封闭标题信息的其它数据来确定最相关动作

的图像中的位置。通常，视频图像中封闭标题信息看起来接近于其言语表现为封闭标题中的文本的一个或几个人。因而，封闭标题信息的坐标通常为最相关动作的图像中的位置提供了良好的预测。

实践中，预测器 22 一般利用视频数据、音频数据以及其它数据的组合来确定图像中的最相关动作位置，并据此建立摇摄及扫描像素坐标。最后，预测器 22 可以应用以下关系式来建立摇摄及扫描像素坐标 P

$$P=X1*F(V)+X2*F(A)+X3*F(D) \quad (1)$$

其中，X1=分配给视频数据的加权因子(0<X1<1)

X2=分配给音频数据的加权因子(0<X2<1)

X3=分配给其它数据的加权因子(0<X3<1)

F(V)=使用视频数据预测的坐标

F(A)=使用音频数据预测的坐标

F(D)=使用其它数据预测的坐标

并且，$X1+X2+X3=1$ (2)

加权因子 X1、X2 和 X3 可以根据经验而定。

预测器 22 的输出 P(即，摇摄及扫描像素坐标)传送到视频处理器 24，该视频处理器 24 可以采用具有对视频图像进行摇摄及扫描格式化操作的性能的编码器、解码器或其它设备的形式。换言之，处理器 24 利用由预测器 22 建立的摇摄及扫描坐标 P 来确定摇摄及扫描图像捕获视窗的位置。建立了摇摄及扫描图像捕获视窗的位置以后，然后处理器 24 可以将视频图像格式化成所希望的宽高比。实践中，处理器 24 可以采取由类似“Media 100 CineStream”软件(可从麻萨诸塞州的 Media 100,Marlboro 处获得)或 MTI IntelliDeck®软件(可从罗得岛州的 Mathematical Techonologies Inc. Providence 处获得)这样的几种公知的视频编辑程序编程的计算机的形式。

实践中，由装置 20 接收的输入视频数据中的帧具有与宽屏视频图像数据相应的 16:9 的宽高比，并且处理器 24 进行摇摄及扫描格式化操作以形成具有 4:3 宽高比的图像。但是，输入视频数据中的帧可以具有其它宽高比，例如在具有标准及额外图像宽度的视频图像衍生电影情况下的 5.5:3 或 7:3。实际上，输入视频图像数据中的帧可以具有 4:3 的宽高比，并且处理器 24 可以根据由预测器 22 预测出的像素坐标来进行摇摄及扫描操作，以便剪辑图像帧来将它们的宽高比改变成为 16:9。

前面描述了这样一种技术，即用于在视频图像帧中建立摇摄及扫描坐标位置，该摇摄及扫描坐标位置定义用于格式化图像以改变其宽高比的图像捕获视窗。

虽然已经针对后期制作动作(即已经被初始捕获过的图像帧的格式化)描述了本发明技术，但技术也可以应用于前期制作动作。实际上，预测器22可以容易地处理“现场”视频图像数据以便在替换宽高比的后续编辑之前产生摇摄及扫描像素坐标。

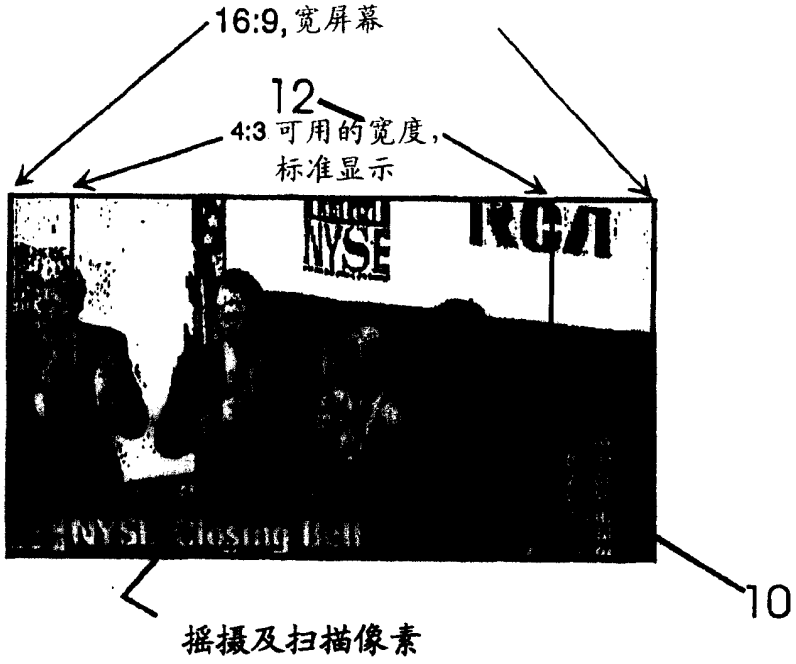

16:9,宽屏幕
12
4:3 可用的宽度，
标准显示
NYSE
RCA
NYSE Closing Bell
10
摇摄及扫描像素

图1

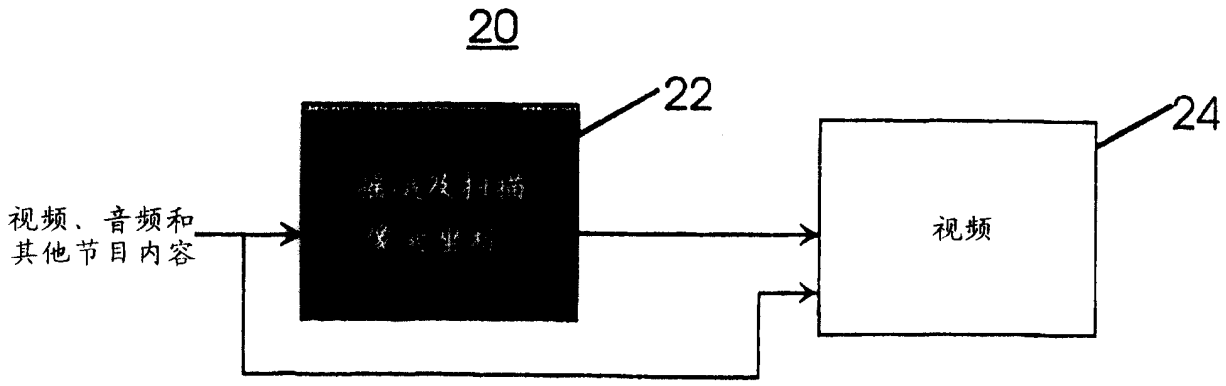

20
22
24
视频、音频和
其他节目内容
视频

图2