

(19) 中华人民共和国国家知识产权局



(12) 发明专利申请

(10) 申请公布号 CN 103581696 A

(43) 申请公布日 2014. 02. 12

(21) 申请号 201310347407. 7

(22) 申请日 2013. 08. 09

(30) 优先权数据

13/571, 479 2012. 08. 10 US

(71) 申请人 英特尔公司

地址 美国加利福尼亚州

(72) 发明人 Y·廖 J·R·福斯特

(74) 专利代理机构 上海专利商标事务所有限公司 31100

代理人 毛力

(51) Int. Cl.

H04N 21/234 (2011. 01)

H04N 21/2343 (2011. 01)

H04N 21/647 (2011. 01)

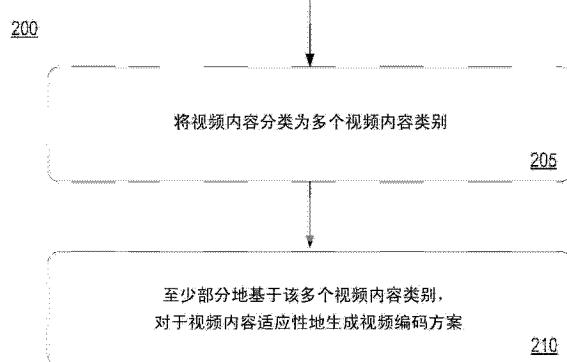
权利要求书2页 说明书6页 附图5页

(54) 发明名称

用于内容感知的多媒体流送的方法与系统

(57) 摘要

一种用于内容感知的多媒体流送的方法与系统，用于将视频内容分类为多个视频内容类别；且至少部分地基于该多个视频内容类别对视频内容适应性地生成视频编码方案。



1. 一种方法,包括 :

将视频内容分类为多个视频内容类别 ;和

至少部分地基于所述多个视频内容类别对所述视频内容适应性地生成视频编码方案。

2. 如权利要求 1 所述的方法,其特征在于,还包括基于所述视频编码方案中的至少一个来生成经编码的视频的输出。

3. 如权利要求 2 所述的方法,其特征在于,还包括 :

确定所生成的经编码的视频输出的视频质量 ;和

基于所确定的视频质量来适应性地生成视频方案。

4. 如权利要求 1 所述的方法,其特征在于,还包括标识所述视频内容的至少一个视频特性并基于至少一个视频特性来分类所述视频内容。

5. 如权利要求 1 所述的方法,其特征在于,所述多个视频内容类别包括代表所述视频内容中不同运动量的至少两个类别。

6. 如权利要求 1 所述的方法,其特征在于,对于所述视频内容适应性地生成视频编码方案还基于如下中至少一个 :视频质量分数、网络条件的指示、用户喜好、以及上述的组合。

7. 如权利要求 1 所述的方法,其特征在于,对于所述视频内容适应性地生成视频编码方案建立了如下参数中的至少一个的值 :目标比特率、编码分辨率、编码帧速率、速率控制算法、帧结构、图片组大小、特定帧类型的数量。

8. 一种系统,包括 :

视频内容分析器,用于将视频内容分类为多个视频内容类别 ;和

内容感知的编码方案生成器,用于至少部分地基于所述多个视频内容类别来对于所述视频内容适应性地生成视频编码方案。

9. 如权利要求 8 所述的系统,其特征在于,还包括视频质量估算模块,用于基于所述视频编码方案中的至少一个来生成经编码的视频的输出。

10. 如权利要求 9 所述的系统,其特征在于,所述视频质量估算模块还确定所生成的经编码的视频输出的视频质量 ;且所述内容感知的编码方案生成器基于所确定的视频质量来适应性地生成视频方案。

11. 如权利要求 8 所述的系统,其特征在于,所述视频内容分析器还标识所述视频内容的至少一个视频特性,且所述内容感知的编码方案生成器基于至少一个视频特性对所述视频内容进行分类。

12. 如权利要求 8 所述的系统,其特征在于,所述多个视频内容类别包括代表所述视频内容中不同运动量的至少两个类别。

13. 如权利要求 8 所述的系统,其特征在于,所述内容感知的编码方案生成器进一步基于如下中至少一个来对于所述视频内容适应性地生成视频编码方案 :视频质量分数、网络条件的指示、用户喜好、以及上述的组合。

14. 如权利要求 8 所述的系统,其特征在于,对于所述视频内容适应性地生成视频编码方案建立了如下参数中的至少一个的值 :目标比特率、编码分辨率、编码帧速率、速率控制算法、帧结构、图片组大小、特定帧类型的数量。

15. 一种其上存储了处理器可执行指令的非瞬态介质,所述介质包括 :

将视频内容分类为多个视频内容类别的指令 ;和

至少部分地基于所述多个视频内容类别对所述视频内容适应性地生成视频编码方案的指令。

16. 如权利要求 15 所述的系统,其特征在于,还包括基于所述视频编码方案中的至少一个来生成经编码的视频的输出的指令。

17. 如权利要求 16 所述的介质,其特征在于,还包括:

确定所生成的经编码的视频输出的视频质量的指令;和

基于所确定的视频质量来适应性地生成视频方案的指令。

18. 如权利要求 15 所述的介质,其特征在于,还包括标识所述视频内容的至少一个视频特性并基于至少一个视频特性来分类所述视频内容的指令。

19. 如权利要求 15 所述的介质,其特征在于,所述多个视频内容类别包括代表所述视频内容中不同运动量的至少两个类别。

20. 如权利要求 15 所述的介质,其特征在于,对于所述视频内容适应性地生成视频编码方案还基于如下中至少一个:视频质量分数、网络条件的指示、用户喜好、以及上述的组合。

21. 如权利要求 15 所述的介质,其特征在于,对于所述视频内容适应性地生成视频编码方案建立了如下参数中的至少一个的值:目标比特率、编码分辨率、编码帧速率、速率控制算法、帧结构、图片组大小、特定帧类型的数量。

用于内容感知的多媒体流送的方法与系统

[0001] 背景

[0002] 网络上的多媒体流送持续以极大速率增长。在一些方面,多媒体流送的持续增长可归于其在新媒体和娱乐应用中不断增加的出现和 / 或重要性、以及在教育、商业、旅游、和其他环境中使用其而获得的增益。在一些情况下,被用于流送多媒体的网络可以是有线或无线的且可包括因特网、电视广播、卫星、蜂窝、和 WiFi 网络。对视频体验而言,重要的是被接收供用户观看的视频的质量。在一些方面,可通过不同优化技术来帮助增加服务能力并增强终端用户的质量体验 (QoE)。

[0003] 在努力增加服务能力和增强终端用户 QoE 方面已经提出了多个适应性视频流送技术。一些这样的技术通过以不同的预先确定的比特率将视频源编码为较短的片段来解决流送能力和质量问题。然后基于可用网络带宽和处理条件在网络上传递视频的经编码的较短的片段。

[0004] 考虑了可用网络带宽和处理条件的技术能解决或无法解决一些宽视频质量项到一定程度,这样的技术没有一般地适应于、响应于、或甚至意识到所传输的视频的各种类型。

[0005] 附图简述

[0006] 本文中所公开的各方面通过示例而非限制地在附图中示出。为说明的简单和清楚而非限制起见,在附图中示出的各方面不一定按比例绘制。此外,在认为合适的地方,在附图中重复附图标记以指示相应或相似的元件。

[0007] 图 1 是与此处视频的一些方面有关的说明性曲线。

[0008] 图 2 是根据此处一个实施例的过程的流程图。

[0009] 图 3 为根据此处的一些实施例的过程的另一个流程图。

[0010] 图 4 是根据实施例的系统的功能性框图。

[0011] 图 5A-5D 是根据此处一些实施例的视频场景的说明性描述。

[0012] 图 6 是根据此处一些实施例的系统的说明性示意框图。

[0013] 详细描述

[0014] 下文描述了通过提供内容感知的视频适应技术,可支持过程与操作来改进视频传输的质量与效率的方法与系统。如下文将更详细地解释地,此处本公开提供了一些实施例,关于基于被编码用于在网络上传输的视频序列的内容来适应性地选择编码参数并分配资源的技术或机制。此处描述的技术(多个)可在一些实施例中,用于最小化比特率消耗和 / 或改进在网路上传输的经编码视频的质量。

[0015] 在一些方面,本公开包括有关用于实现此处的过程与系统的方法(多个)和系统(多个)的特定细节。然而,本领域技术人员将领会,没有这些具体细节也可实践本公开内容的诸实施例。因而,在一些实例中,诸如控制机制和全软件指令序列等各方面未被详细示出以便不模糊本公开内容的其他方面。本领域技术人员利用本文所包括的描述将能在无需过度实验的情况下实现适当的功能。

[0016] 在本公开中对“一个实施例”、“一些实施例”、“一实施例”、“示例实施例”、“一实

例”、“一些实例”等的引用表明所描述的实施例可包括特定特征、结构或特性，但不一定每个实施例均包括该特定特征、结构或特性。此外，这样的短语不一定是指同一个实施例。此外，当结合实施例描述特定特征、结构或特性时，认为本领域技术人员知道结合无论是否明显描述的其他实施例来实现这些特征、结构或特性。

[0017] 本文的一些实施例可以用硬件、固件、软件或它们的任意组合来实现。各实施例还可实现为储存在机器可读介质上的可执行指令，这些指令可由一个或多个处理器读取和执行。一种机器可读存储介质可包括用于以机器(例如，计算设备)可读的形式存储信息的任意有形的非暂时性机构。在一些方面，机器可读存储介质可包括只读存储器(ROM)；随机存取存储器(RAM)；磁盘存储介质；光存储介质；闪存设备；以及电子或光学形式的信号。尽管在本文中将固件、软件、例程、以及指令描述为执行某些动作，但应当明白，这些描述仅仅是出于方便起见且这些动作实际上得自执行该固件、软件、例程以及指令的计算设备、处理器、控制器以及其他设备。

[0018] 图 1 是示出在不同编码设置下各种视频内容的所观察到的速率质量特性的曲线 100。在曲线 100 中反映的视频的内容特性各有不同。例如，某些视频内容可包括非常少的动作(如，坐在桌边的新闻节目主持人的新闻广播)而某些视频可包括大量动作(如，有很多选手绕比赛场地同时移动的体育赛事)。例如，编码设置可包括帧结构、GOP(图片组)大小、等。参看曲线 100，水平轴 105 表示比特率刻度且垂直轴 110 代表视频质量的度量(即，多刻度结构性类似性(MS-SSIM)指数)刻度。曲线 100 示出这一点：对于在相同比特率处被编码的视频，视频质量可在较大范围内变化。例如，在 4Mbps 速率处，对于在不同设置处被编码的不同视频，平均 MS-SSIM 值从约 0.87 变化到约 0.98。还有，曲线 100 展示出，例如，对于 0.95 的 MS-SSIM，所需要的比特率可从约 2Mbps 变化至约 14Mbps。

[0019] 因此，曲线 100 展示出，对于所有视频内容使用固定编码参数(多个)的视频编码与传输方法可导致带宽的浪费或视频质量的劣化。

[0020] 图 2 为根据此处的实施例的过程 200 的说明性流程图。过程 200 可通过确定最小化比特率消耗同时还满足用户 QoE 标准的最优、或至少更有效的编码方案(profile)，以将源自不同视频内容的速率质量性能的较大变化考虑在内。

[0021] 在操作 205 处，输入的视频内容可被分为各种视频内容类别。在操作 205 处接收到的视频可来自任何源，包括现场反馈，且可从存储位置被取回。可基于视频本身(即，视频的内容)的一个或多个特性来分类在操作 205 处接收到的视频。在一些实施例中，所接收到的视频的动作密度特性可被评估且该视频可被分类为三个类别中的一个：低动作量、中等动作量、或高动作量。

[0022] 在操作 210 处，可至少部分地基于在操作 205 处被确定的多个视频内容来对该视频内容适应性地生成一个或多个视频编码方案。如图 2 中所示，操作 210 可接收来自操作 205 的视频内容类别的指示。在一些方面(如下文中详细讨论地)，操作 210 可接收附加信息作为除来自操作 205 的视频内容类别信息外的输入。操作 210 可使用来自操作 205 的视频内容类别和其他信息来针对不同类别的视频内容适应性地生成编码方案。要注意的是不同类别的视频内容各自相关或关联于不同类型的视频内容(即，具有不同特性的视频)。

[0023] 至少部分地基于所确定的多个视频内容类别的在操作 210 处适应性地生成的编码方案可被存储或输出在记录或文件内，被用作输入来进一步处理并传输视频内容、以及

用于其他过程。

[0024] 图 3 涉及根据此处的一些实施例的处理器 300。在一些方面，处理器 300 类似于图 2 的处理器 200。例如，操作 305 和 310 可分别对应于操作 205 和 210。因此，此处不提供操作 305 和 310 的详细讨论，因为通过参加上述操作 205 和 210 的讨论可对于这些操作有完整的理解。

[0025] 参看图 3，基于在操作 310 处适应性地生成的至少一个视频编码方案，操作 315 生成经编码的视频的输出。在操作 320，可使用操作 315 的输出来确定或计算经编码的视频的视频质量分数或测量。在操作 320 处确定的视频质量分数可提供对于经编码的视频的质量的指示。在一些方面，视频质量分数可包括根据一个或多个 VQA 算法而计算的视频数量估算(VQA)度量。

[0026] 如图 3 中进一步所示地，在操作 320 处确定的视频质量分数可被传输至操作 310，从而在操作 310 处所用于生成编码方案的编码参数可被递归地调整从而部分地基于视频内容类别和经编码的视频内容的质量来适应性地生成编码方案。

[0027] 图 4 是根据此处一些实施例的，装置或设备 400 的功能性框图的说明性示图。在一些方面，设备 400 可包括内容感知的多媒体流送服务器来实现此处公开的过程(如，过程 200 和 300)的一些部分。在一些实施例中，设备 400 可以硬件、固件、或其组合实现。在一些方面，设备 400 可包括毕图 4 中具体示出的更多、更少、类似、或可选的功能性组件。在一些实施例中，图 4 中所示的功能性框图可被实现为一个或多个组件、以及与其他功能和 / 或组件组合。

[0028] 视频内容由视频源 405 提供或从视频源 405 接收。视频源 405 可以是用于提供视频内容的任何类型的机制，包括现场或重播数据流以及包括从存储设施(即，存储器)取回的视频序列的文件或记录。来自视频源 405 的视频内容被馈入视频内容分析器 410。视频内容分析器 410 可用于分析来自视频源 405 的视频的内容特性。在一些实施例中，视频内容分析器 410 可包括视频特征提取机制或技术来标识视频内容的不同特性。视频内容分析器 410 可基于被分类的视频内容进一步将视频内容分为不同类别(如，操作 205 和 305)。

[0029] 与由视频内容分析器 410 所分析的视频内容相关联的不同视频类别的指示被提供至内容感知的编码方案生成器 415。内容感知的编码方案生成器 415 可从多个源收集信息以针对不同类型的视频内容适应性地生成最优编码方案。在一些实施例中，不同类型的视频内容对应于不同类别的视频内容。在一些方面，内容感知的编码方案生成器 415 输入的输入信息可包括，至少，来自视频内容分析器 410 的视频内容类别。内容感知的编码方案生成器 415 的附加输入内容可包括，例如，在服务器 400 处由视频质量估算工具 430 所计算的视频质量分数、以及网络条件、以及其他用户要求反馈 420。

[0030] 编码方案生成器 415 可用于通过单独地或组合地适应性地选择目标比特率、编码分辨率、编码帧速率、速率控制算法、帧结构、图片组(GOP)大小、特定类型帧(如，双向“B”帧)的数量、和其他编码参数，来生成一个或多个内容最优的编码方案。将理解的是，本公开包含这些和其他编码参数，无论是否特定地在此处被枚举。

[0031] 编码方案生成器 415 可向多媒体流送编码译码器 425 提供藉此生成的一个或多个内容最优的编码方案。编码译码器 425 可使用内容最优的编码方案来用对来自视频源 405 的视频内容用由视频编码方案生成器 415 所生成的合适的编码方案进行编码。合适的编码

方案(多个)可最优化地匹配视频中内容的类型。

[0032] 由编码译码器 425 所输出的经编码的视频,被部分地向视频质量估算(VQA)工具 430 提供。VQA 工具 430 可对于该经编码的视频来计算视频质量或 VQA 分数(多个)。可将该 VQA 分数(多个)传输至内容感知的编码方案生成器 415。一旦接收到 VQA 分数,内容感知的编码方案生成器 415 可递归地调节此处使用的编码参数并至少部分地基于该视频内容和该 VQA 分数来生成最优的编码方案。

[0033] 在一些实施例中,可使用诸如 MS-SSIM 之类的基于参考的 VQA 分数,因为视频源在服务器侧是可用的。

[0034] 通过使用此处公开的内容感知视频适应过程确定比特率最小化并将它们与对于所有视频序列使用固定的编码方案的基线编码计划(scheme)进行比较,申请人已经了解此处公开的过程的有效性。在评估和如下表格中使用的视频序列包括公开各种比特率下的可获得的“Aspen”、“ControlledBurn”、“RedKayak”、“SpeedBag”、“TouchdownPass”、和“WestWindEasy”视频序列。

[0035] 下面的表 1 示出与其中对于所有输入视频序列应用固定编码方案的基线计划相比较,对于内容感知的视频适应方法(多个)所观察到的增益。在表 1 的示例中,假设当平均 PSNR(峰值信号噪声比)大于 34dB 时用户满意。与表 1 相关的基线计划使用固定的量化参数(QP)来编码视频序列,而内容感知(即,最优的)方法基于在输入视频序列中检测到的视频内容特性的不同类型来适应性地选择编码参数。如所示,表 1 中列出的结果示出,为了使得用户对于所有视频序列满意,使用此处公开的内容感知的视频适应过程获得了 3.55Mbps 的平均比特率的节省。

[0036]

序列	基线 (QP = 34)		基线 (QP=32)		最优	
	比特率 (Mbps)	平均 PSNR (dB)	比特率 (Mbps)	平均 PSNR (dB)	比特率 (Mbps)	平均 PSNR (dB)
Aspen	7.94	34.74	10.03	35.80	4.89	34.17
Controlledburn	6.45	33.65	8.07	34.75	4.90	34.03
Redkayak	8.01	34.00	10.14	35.14	7.65	34.11
Speedbag	6.44	39.02	7.58	39.81	2.12	35.62
Touchdownpass	4.02	36.04	5.01	36.76	2.18	34.02
Westwindeasy	7.26	33.56	9.12	34.77	6.92	34.26
比特率/用户 满意度	6.69	66.7%	8.32	100%	4.77	100%

[0037] 表 1

[0038] 作为示例,下文的表 2 提供了对于表 1 的每一个视频序列的编码参数设置的列表。

[0039]

序列	速率控制	GOP 大小	B 帧数量
Aspen	VBR=5Mbps	30	2
Controlledburn	QP=32, Δ P / Δ B=2	15	2
Redkayak	QP=32, Δ P / Δ B=2	15	0
Speedbag	CBR=2Mbps	30	0
Touchdownpass	QP=38, Δ P / Δ B=2	30	0
Westwinedeasy	QP=30, Δ P / Δ B=2	30	2

[0040]

[0041] 表 2

[0042] 图 5A–5D 绘图般地示出此处公开的视频内容的适应性编码分辨率的过程可如何改进视频序列的视频质量的示例。视频序列“Controlledburn”(图 5A 和 5B) 以及“Redkayak”(图 5C 和 5D) 各自在 220x124 分辨率(图 5A 和 5C) 和 768x432 分辨率(图 5B 和 5D) 被编码。注意的是这些视频序列均在相同比特率(即, 230kbps) 被编码。对于“Controlledburn”视频序列, 如图 5B 中所示以更高的分辨率来编码减少了视频的模糊并改进了所知觉的视频质量。然而, 如图 5D 中所示, 以更高的分辨率来对“Redkayak”视频序列进行编码导致视频看起来非常斑驳且劣化了视频质量。因此, 证明了, 对于特定类型(多个)(即, 视频特定)的视频序列的视频内容的适应性的编码参数(如, 编码分辨率等)可有效地增强视频流送服务、应用、系统、过程、或设备的 QoE。

[0043] 图 6 是根据一些实施例的系统或装置 600 的框图概览。系统 600 可例如与实现此处描述的方法与过程的任何设备相关联, 这些设备包括例如供应多媒体数据的流送服务提供器的服务器(如, 图 4, 设备 400)、或任何其他实体系统 600 包括处理器 605, 诸如, 例如, 形式为单芯片微处理器或多核处理器的一个或多个商业可用的中央处理单元, 耦合至通信设备 615, 被配置为经由通信网络(图 6 中未示出) 与另一个设备或系统进行通信。在该实现中, 系统 600 包括应用服务器, 通信设备 615 可提供系统 600 来与客户机设备交互。系统 600 还可包括本地存储器 610, 诸如 RAM 存储器模块。系统 600 还包括输入设备 620(如, 用于输入内容的触摸屏、鼠标、和 / 或键盘) 和输出设备 625(如, 用于显示用户界面的计算机或其他设备监视器 / 屏幕)。

[0044] 处理器 605 与存储设备 630 通信。存储设备 630 可包括任何合适的信息存储设备, 包括磁存储设备(如, 硬盘驱动器)、光存储设备、和 / 或固态存储器设备的半导体的组合。在一些实施例中, 存储设备可包括数据库系统。

[0045] 存储设备 630 存储程序代码 635, 可提供计算机可执行指令用于, 根据此处的过程而处理来自例如客户机设备的处理请求。处理器 605 可执行程序 635 的指令来藉此根据此处描述的任何实施例地进行操作。程序代码 635 可被存储为压缩的、未编译的、和 / 或加密

的格式。程序代码 635 可进一步包括其他程序元件，诸如操作系统、数据库管理系统、和 / 或由处理器 605 使用来与例如外围设备交互的设备驱动器。存储器设备 630 还可包括数据 645，诸如是视频序列和 / 或用户喜好或设置。在一些方面，数据 645 与内容感知的编码方案生成器 640 结合，可由系统 600 用于执行此处的过程，诸如过程 200 和 300。

[0046] 本文讨论的所有系统和过程可被实现在储存于一个或多个计算机可读介质上的程序代码中。这样的介质可包括例如软盘、CD-ROM、DVD-ROM、一种或多种类型的“盘”、磁带、存储卡、闪存驱动器、固态驱动器、以及固态随机存取存储器(RAM)、只读存储器(ROM)存储单元、和其他非瞬态介质。进一步，此处公开或引用的系统和装置可包括硬件、软件、和固件，包括通用目的、专用、和分布式计算设备、处理器、处理核、和微处理器。在一些方面，此处公开的过程与方法可被传递并提供为服务。各实施例因而不限于硬件和软件的任何具体组合。

[0047] 本文中仅仅出于说明目的描述了各实施例。本领域技术人员将从这一描述中认识到，各实施例不限于所描述的实施例，而是可以在只受权利要求书的精神和范围所限定的各种修改和更改的情况下实施。

100

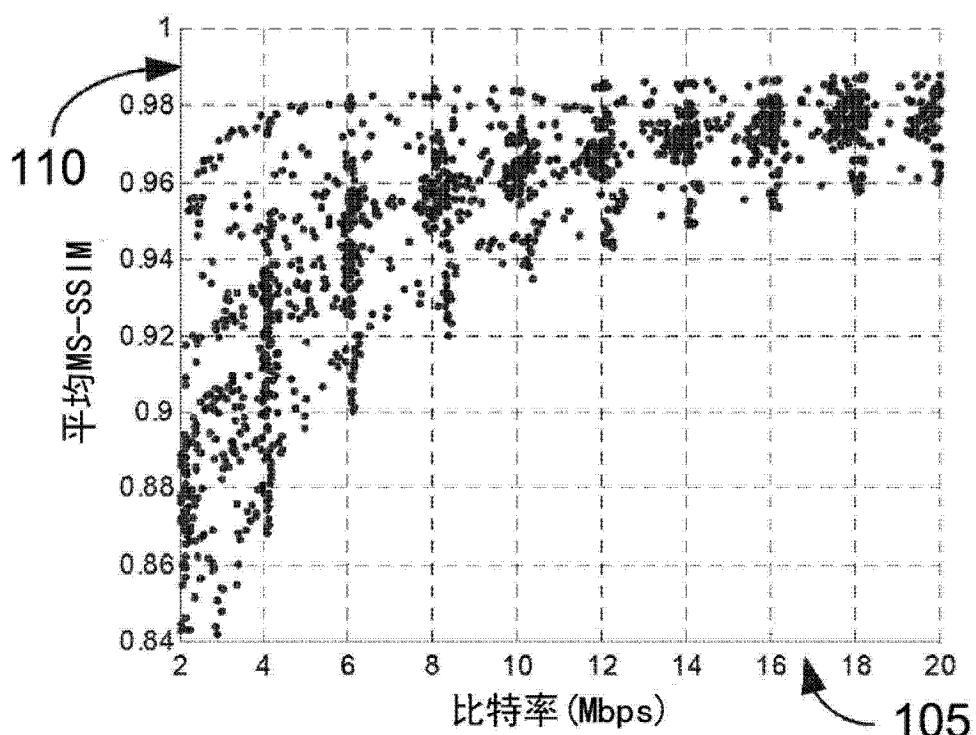


图 1

200

将视频内容分类为多个视频内容类别

205

至少部分地基于该多个视频内容类别，
对于视频内容适应性地生成视频编码方案

210

图 2

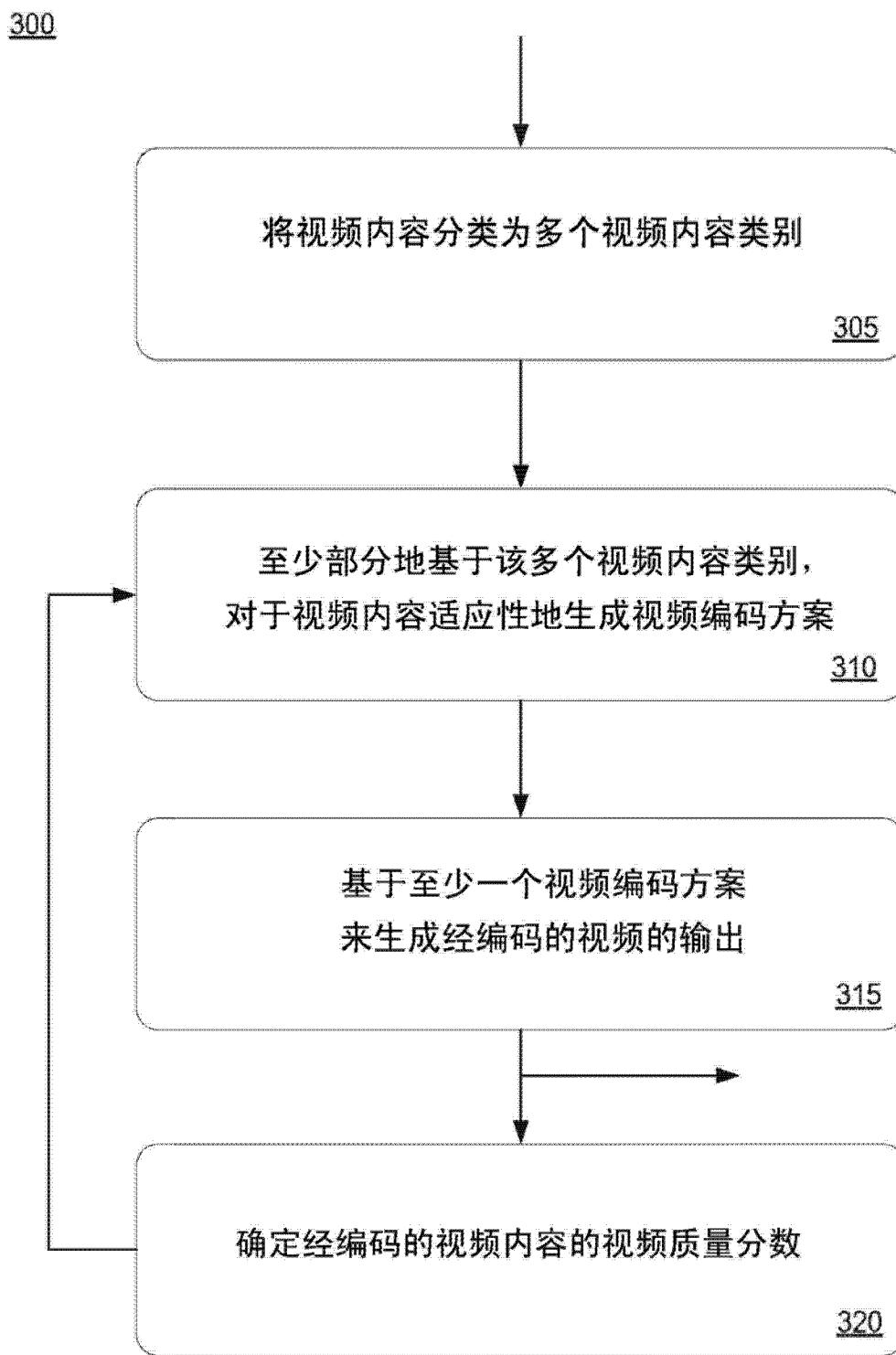
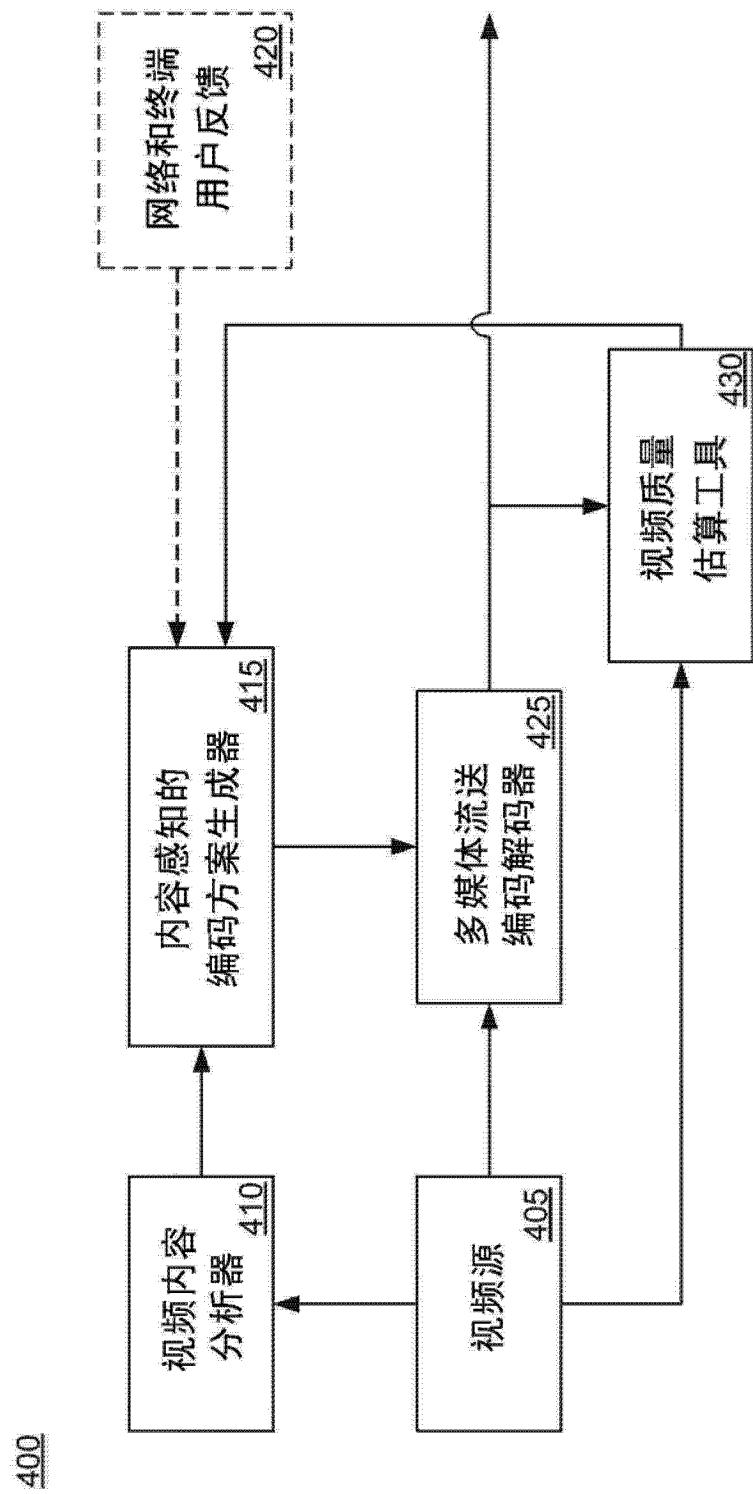


图 3



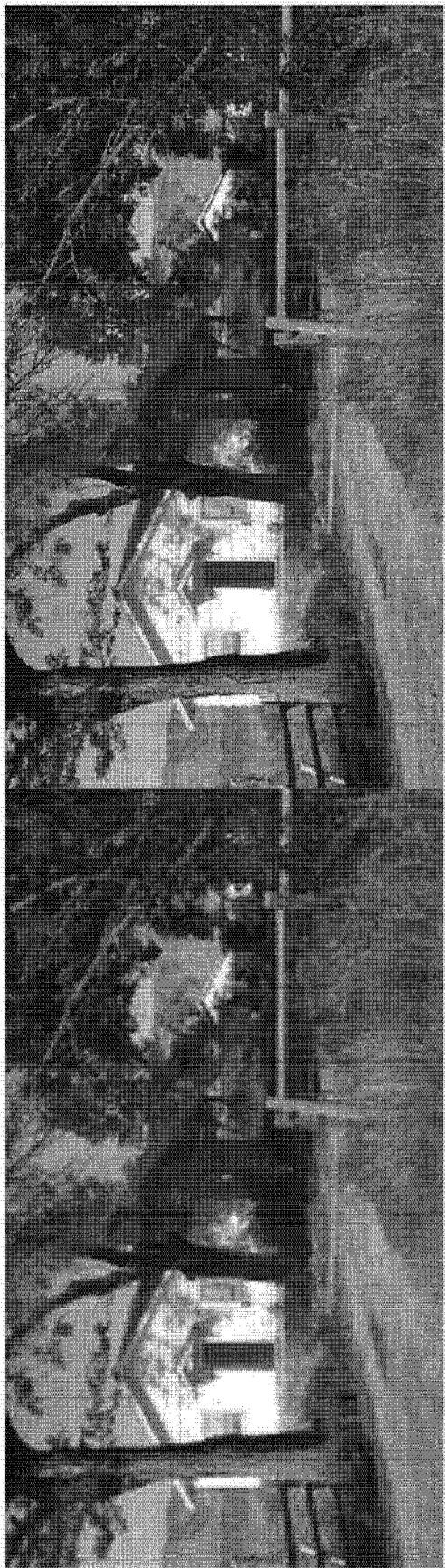


图 5A



图 5B



图 5C



图 5D

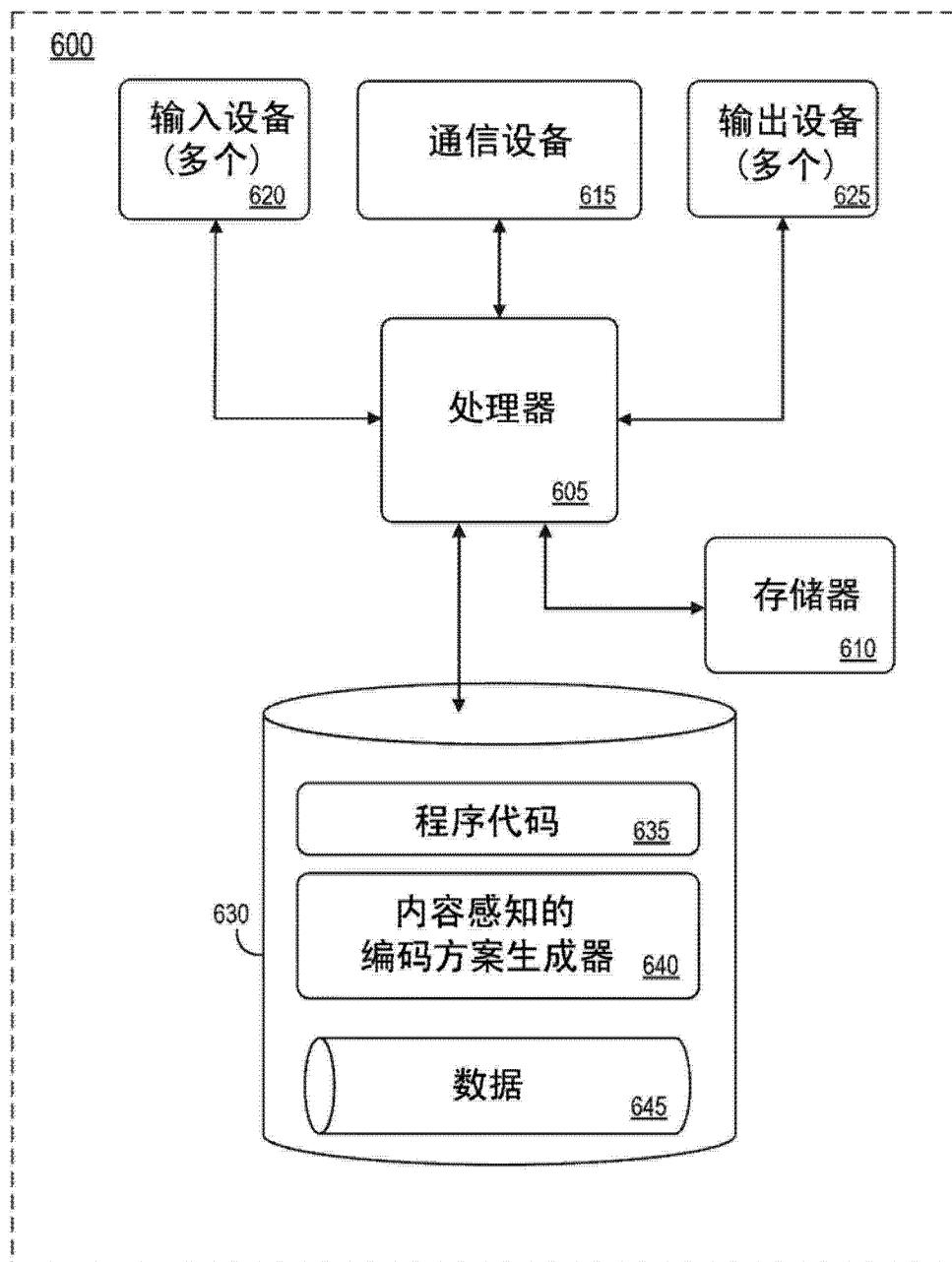


图 6