



# (12) 发明专利申请

(10) 申请公布号 CN 102884536 A

(43) 申请公布日 2013. 01. 16

(21) 申请号 201080066641. 7

(51) Int. Cl.

(22) 申请日 2010. 09. 22

G06K 9/00 (2006. 01)

(30) 优先权数据

H04N 7/14 (2006. 01)

12/755, 551 2010. 04. 07 US

H04N 7/26 (2006. 01)

(85) PCT申请进入国家阶段日

2012. 11. 07

(86) PCT申请的申请数据

PCT/US2010/049828 2010. 09. 22

(87) PCT申请的公布数据

W02011/126500 EN 2011. 10. 13

(71) 申请人 苹果公司

地址 美国加利福尼亚

(72) 发明人 F·多普克

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 邹姗姗

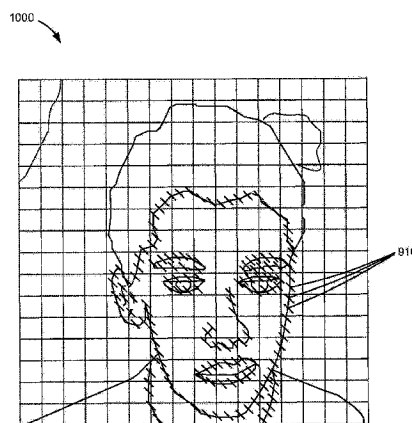
权利要求书 2 页 说明书 9 页 附图 12 页

## (54) 发明名称

用于视频会议压缩的肤色与特征检测

## (57) 摘要

在许多视频会议应用中,带宽是非常珍贵的,因而,智能地编码给定的视频帧是很重要的。常常期望更大量的信息花费在编码视频帧较重要的部分上,例如,人的脸部特征,而视频帧不太重要的部分可以用较高的比率压缩。因而,需要可以向视频编码器“建议”智能宏块压缩比的用于智能肤色与脸部特征识别视频会议压缩的装置、计算机可读介质、处理器与方法。对压缩比的建议可以至少部分地基于对给定视频帧中哪些宏块有可能包含肤色、有可能包含特征(例如,边缘)、有可能包括在肤色区域中或者其附近的特征、或者有可能既不包含肤色又不包含特征的确定。



1. 一种处理器,编程为执行视频压缩方法,该方法包括:  
在处理器处接收视频数据,所述视频数据具有带多个区域的帧,每个区域具有多个像素;  
利用处理器确定每个区域中具有预定色调的任何第一像素;  
利用处理器确定每个区域中作为至少一个特征的一部分的任何第二像素;  
利用处理器基于对视频帧的每个区域中任何第一像素和第二像素的确定给该区域计分;及  
利用处理器基于所述计分压缩视频帧的区域。
2. 如权利要求 1 所述的处理器,其中,确定第一像素包括将区域中像素的第一值与所述预定色调进行比较。
3. 如权利要求 2 所述的处理器,其中,所述预定色调包括指示人肤色的 CbCr 值。
4. 如权利要求 2 所述的处理器,其中,所述预定色调包括指示人肤色的 RGB 值。
5. 如权利要求 1 所述的处理器,其中,确定第二像素包括对区域中的像素执行边缘检测处理。
6. 如权利要求 1 所述的处理器,其中,确定第二像素包括对区域中的像素执行特征检测处理。
7. 如权利要求 1 所述的处理器,其中,给每个区域计分包括:如果每个区域的给定像素具有所述预定色调,则针对所述给定像素分配第一得分。
8. 如权利要求 7 所述的处理器,其中,给每个区域计分包括:如果每个区域的给定像素是至少一个特征的一部分,则针对所述给定像素分配第二得分。
9. 如权利要求 8 所述的处理器,其中,所述第二得分高于所述第一得分。
10. 如权利要求 8 所述的处理器,其中,给每个区域计分包括:如果每个区域的给定像素是至少一个特征的一部分而且在具有所述预定色调的像素的阈值距离之内,则针对所述给定像素分配第三得分。
11. 如权利要求 10 所述的处理器,其中,所述第三得分大于所述第二得分,而且其中所述第二得分大于所述第一得分。
12. 如权利要求 10 所述的处理器,其中,给每个区域计分包括:对分配给每个区域中的像素的值求平均值;并且,将该平均值与所述区域相关联。
13. 如权利要求 1 所述的处理器,其中,压缩区域包括:利用比具有较低得分的区域少的压缩来压缩具有较高得分的区域。
14. 如权利要求 1 所述的处理器,其中,压缩区域包括:  
将一个给定区域的得分与一个或多个邻近区域的得分进行比较;及  
基于差异调整给定区域的得分。
15. 如权利要求 14 所述的处理器,其中,所述差异指示,与所述一个或多个邻近区域相比,所述给定区域具有更低或更高的得分。
16. 如权利要求 1 所述的处理器,其中,每个区域包括一个宏块。
17. 一种视频压缩方法,包括:  
接收视频数据,该视频数据具有带多个区域的帧,每个区域具有多个像素;  
确定每个区域中的颜色值在预定色调区域内的任何第一像素;

- 确定每个区域中作为至少一个特征的一部分的任何第二像素；  
对于每个区域中的每个像素：  
如果该像素是第一像素但不是第二像素，则针对该像素分配第一值；  
如果该像素是第二像素但不是第一像素，则针对该像素分配第二值；  
如果该像素既不是第一像素也不是第二像素，则针对该像素分配第三值；及  
如果该像素是第二像素并且是第一像素或者在第一像素的阈值距离内，则针对该像素分配第四值；  
基于视频帧的每个区域中像素的平均分配值，给该区域计分；及  
基于视频帧的每个区域的得分来压缩该区域。
18. 如权利要求 17 所述的方法，其中，所述第四值大于所述第一值、第二值和第三值。
19. 如权利要求 18 所述的方法，其中，所述第一值和第二值各自都大于所述第三值。
20. 一种视频压缩方法，包括：  
接收视频数据，该视频数据具有带多个区域的帧，每个区域具有多个像素；  
确定每个区域中的颜色值在预定色调区域内的任何第一像素；  
确定每个区域中作为至少一个特征的一部分的任何第二像素；  
基于对每个区域中任何第一像素和第二像素的确定给该区域计分；  
将得分与一个或多个邻近区域的得分相差多于阈值的每个区域的得分校正成等于邻近区域的平均得分；及  
基于视频帧的每个区域的得分来压缩该区域。
21. 如权利要求 20 所述的方法，其中，所述预定色调区域包括指示人肤色的 CbCr 值。
22. 如权利要求 20 所述的方法，其中，确定第二像素包括对区域中的像素执行特征检测处理。
23. 一种装置，包括：  
图像传感器，用于获得视频数据；  
存储器，操作地耦接到所述图像传感器；及  
处理器，操作地耦接到所述存储器与图像传感器，并且编程为编码视频数据，该处理器配置成：  
接收视频数据，该视频数据具有带多个区域的帧，每个区域具有多个像素；  
确定每个区域中具有预定色调的任何第一像素；  
确定每个区域中作为至少一个特征的一部分的任何第二像素；  
至少部分地基于对视频帧的每个区域中任何第一像素和第二像素的确定给该区域计分；及  
至少部分地基于所述计分压缩视频帧的区域。
24. 如权利要求 23 所述的装置，其中该装置包括以下至少一种：数码照相机、数码摄像机、移动电话、个人数据助理、便携式音乐播放器和计算机。
25. 一种其中包含了计算机可读程序代码的计算机可用介质，其中所述计算机可读程序代码适于执行成实现由权利要求 1 所述的编程处理器执行的方法。

## 用于视频会议压缩的肤色与特征检测

[0001] 由于 Apple 公司的 iPhone 4 原型机很显然被 Apple 公司的工程师于 2010 年 3 月 25 日窃取,因此本申请中公开并请求保护的发明过早地并且在没有 Apple 公司授权的情况下公布于众。在上述明显窃取之前尚未提交本申请所基于的美国优先权申请。

[0002] 对相关申请的交叉引用

[0003] 本申请是基于 2010 年 4 月 7 日提交的美国专利申请第 12/755,551 号的国际申请,所述美国申请通过引用并入于此并且要求其优先权。

### 背景技术

[0004] 在许多视频会议应用中,带宽是非常珍贵的,因而,尽可能有效并智能地编码给定的视频帧及整个视频流是很重要的。视频压缩算法一般对数据的宏块操作,宏块即视频帧内相邻像素的方形组。宏块的大小一般是 16 像素乘 16 像素,但是,依赖于编码视频数据所使用的编解码器,例如,H. 263 或者 H. 264,帧也可以分成更小的宏块,例如一边尺寸为 4、8、12 或者 16 像素的宏块。当然,视频帧还可以分成任意尺寸的更小或更大的宏块。

[0005] 在视频会议应用中,在最低可能的比特率情况下确保最高可能的视频质量是极为重要的。当今现有的视频编解码器标准既采用帧间编码又采用帧内编码技术。帧内编码技术是关于只包含在当前视频帧中的信息而不关于视频序列中的任何其它帧来执行的。帧内编码利用人眼不会像察觉到亮度变化那么容易地察觉到非常细小的颜色不同的事实。另一方面,帧间编码技术涉及时间处理,即,不是为后续的视频帧重新发送全部信息,编解码器将仅编码和发送从一个视频帧到下一个视频帧的像素位置与像素值的变化。这是一种有效的技术,因为大量像素常常在从一个视频帧到下一个时将不会变化,因而对每个视频帧重新发送全部的图像信息将是冗余的。

[0006] 在编码过程中,视频压缩编解码器必须做出关于针对每个视频帧在什么地方“花费”其有限数量的比特的关键性决定,即,编解码器必须确定在哪些宏块上需要最大量的图像细节。常常期望更大量的信息用于编码视频帧最重要的部分,而视频帧不太重要的部分可以用较高比率压缩,而且仍然产生令人满意的视频质量。尤其是在视频会议应用中,视频帧最重要的部分通常是人的脸或手在视频帧中的定位,而且,更具体而言,是通信中所涉及的面部特征(例如,眼睛和嘴巴)的定位。

[0007] 因而,需要可以向视频编码器“建议”宏块压缩比的用于智能肤色与面部特征识别视频会议压缩的装置、计算机可读介质、处理器与方法。对压缩比的建议可以至少部分地基于给定视频帧中哪些宏块有可能是最重要的,并因此值得在其编码中花费比在视频帧中不太重要宏块的编码中不成比例地更大量的比特。

### 发明内容

[0008] 现有技术的视频会议压缩方法公开了,通过检查给定视频帧中的每个宏块并确定该宏块中属于预定肤色颜色区域内的像素个数,来定位所述视频帧中的脸部。如果宏块包含高于阈值个数的肤色像素,那么该宏块可以用较低的比率压缩,因为它有可能包含重要

信息,例如,人脸。相反,如果给定的宏块不包含阈值个数的肤色像素,那么现有技术的视频会议压缩方法就把该宏块视为不太重要的并且用较高的比率来压缩它。仅仅基于肤色检测的压缩方案可能导致不期望的视频流,其中,当脸的外边缘只覆盖给定宏块的一小部分时,人脸的外边缘被高度压缩(因而缺乏质量),由此造成算法认为该宏块没有“足够的肤色”,因而不值得用较低的压缩比。另外,现有技术的视频会议压缩方法将有可能花费不必要的大量信息去编码视频帧中并不真正拥有很多细节或特征的肤色区域,例如,前额和脸颊。

[0009] 本公开内容涉及既考虑肤色又考虑特征检测的用于更智能视频压缩的装置、计算机可读介质、处理器与方法。例如,在一种实施方式中,赋予宏块中的每个像素一个加权值,这可以基于像素:a)在肤色区域内(例如,值为“1”);b)是“特征”的一部分(例如,值为“2”);c)是“特征”的一部分而且在肤色区域的阈值距离(例如,两个像素)内内(例如,值为“16”)或者d)既不在肤色区域内也不是“特征”的一部分(例如,值为“0”)。上述0:1:2:16加权方案仅仅是示例性的,只是例示与只包含肤色特征或者只包含特征的宏块相比,出现在肤色区域中或者在其附近的特征可以被认为对于视频帧的编码呈指数倍地更加重要。确定给定像素是否为“特征”的一部分可以基于任意数量的众所周知的特征检测算法,例如 Viola Jones 对象检测框架,或者任意数量的众所周知的边缘检测算法,例如 Sobel 边缘检测。

[0010] 然后,可以计算用于给定宏块的加权像素值的平均值,而且,可以基于计算出的平均加权像素值赋予宏块整体“压缩得分”。采用加权平均值允许比较不同尺寸的宏块的相对重要性。压缩得分可以基于例如1至10的数值范围,其中具有压缩得分10的宏块被压缩得最少,而具有得分1的宏块被压缩得最多。因而,在此公开的智能视频压缩方案朝着对视频会议最重要的区域(即,脸、手,而且更特别地是脸部特征和脸的外边缘)优化视频帧上宏块压缩比的分布。尽管本公开内容讨论了用于视频会议应用中更智能视频压缩的一种新技术,但是,本领域的普通技术人员将认识到,在此所公开的技术还可以应用到其它的背景和应用中。

[0011] 以上内容并非旨在概述本公开的每种可能的实施方式或者每个方面。

#### 附图说明

[0012] 图1至2图示根据本公开一种实施方式的多媒体设备。

[0013] 图3示出了根据本公开一种实施方式的基于肤色和特征检测来压缩视频的处理。

[0014] 图4示出了根据一种实施方式的分成多个宏块的视频帧。

[0015] 图5至6示出了根据一种实施方式且更具体的用于基于肤色和特征检测来压缩视频的处理。

[0016] 图7至8例示了根据一种实施方式的在色度空间中的预定“肤色区域”。

[0017] 图9示出了根据一种实施方式的对具有各种像素的一组宏块运行的假设特征检测算法的结果。

[0018] 图10示出了根据一种实施方式的作为特征的一部分而且还在肤色区域的阈值距离内的假设图像的区域。

[0019] 图11示出了根据一种实施方式的用于调整处理过的帧的压缩得分的处理。

[0020] 图12示出了根据一种实施方式的具有宏块离群值(outlier)的视频帧。

## 具体实施方式

[0021] 图 1 中图示出的多媒体设备 100 具有编程为根据本公开执行用于对视频数据计分和编码的指令的处理单元 120。在这种实施方式中,多媒体设备 100 具有在视频会议中处理音频数据的音频编解码器 130、麦克风 132 和扬声器 134。同样,多媒体设备 100 具有在视频会议中处理视频数据的视频编解码器 140、图像传感器 142 和显示器 144。最后,多媒体设备 100 还可以包括其它部件,包括用户接口 112、通信电路 114、存储器 116 及本领域中已知的其它部件。存储器 116 也是计算机可读介质的一个例子,用于执行智能视频压缩方法的指令可以存储(即,有形地体现)在其中,而且它可以包括 ROM、EEPROM、闪速存储器或者任何其它合适类型的存储器存储介质。

[0022] 在此所公开的技术可以适用于任意数量的具有数字图像传感器的消费者电子设备。例如,多媒体设备 100 可以是数码照相机、数码摄像机、移动电话、个人数据助理(PDA)、便携式音乐播放器、计算机、视频会议端点或者其它设备。对于本公开而言,设备 100 可以称为移动电话。由此,所公开的视频压缩技术对于具有前置摄像头的电话会是有用的,因而,近端讲话者的脸可以准确地呈现在远端参与者的电话的前显示器上,反之亦然。

[0023] 图像传感器 142 可以是包括用于拍摄视频(和静止图像)的电荷耦合器件(CCD)的传统照相机图像传感器,而且可以集成到多媒体设备 100 中。处理单元 120 可以是嵌入式处理器,例如具有 **ARM®** v7-A 架构的 **Cortex®** A8,该处理器为执行所公开的技术提供了多用途而且鲁棒的计算单元。(CORTEX 和 ARM 是英国 ARM 有限公司的注册商标。)

[0024] 视频编解码器 140 接收从图像传感器 142 输出的视频数据并且压缩该视频数据,用以经通信电路 114 在通信网络上发送。为此,视频编解码器 140 压缩视频数据,使得它可以根据网络的带宽限制在网络上发送。

[0025] 给定以上对多媒体设备 100 (图 1) 的综述,现在讨论设备的附加细节。图 2 示出了例示用于图 1 的多媒体设备 100 的处理单元的图像传感器组件(Image Sensor Package, ISP) 200 和处理部件 220 的更多细节的框图。为了捕捉视频数据,ISP 200 具有数字图像传感器 202,该传感器 202 可以是 CCD 等。依赖于所使用的传感器的类型,视频数据可以通过去马赛克处理或者去 Bayer 处理(de-Bayering process) 204 来处置,如本领域中众所周知的。而且,可以执行各种校正处理 206。例如,可以进行透镜校正,来校正各种透镜变形的视频数据,包括晕映假象,即,光朝着帧的边缘衰减,和颜色一致性。本领域中已知的 RGB 对比度与饱和度提升可以偏移视频数据的中点的颜色范围,而降噪滤波器可以除去由低光条件等造成的任何噪声。最后,ISP 200 可以把视频数据编码到 YCbCr 色彩空间中,并且经接口 210 传递编码后的视频数据,用于进一步的处理。

[0026] 如已知的,YCbCr 色彩空间中的像素具有用于照度或亮度的 Y 分量、用于“蓝色差值”色度分量(即,B-Y)的 Cb 分量和用于“红色差值”色度分量(即,R-Y)的 Cr 分量。Cb 和 Cr 分量一起称为“色度空间”。用于多媒体设备的许多数码照相机和数码摄像机都提供二次采样的 YCbCr 色彩空间中的图像数据,以节约存储器带宽。例如,利用所谓的 4:2:0 二次采样,Cb 和 Cr 是以 Y 的一半垂直分辨率和一半水平分辨率采样的,使得四分之一数量的色度样本作为亮度样本用于图像的给定区域。尽管视频数据的处理在这里是使用这种 YCbCr 色彩空间讨论的,但是本公开的教习可以用于任何合适的色彩空间与采样。

[0027] 从接口 210, 编码后的视频数据随后由多媒体设备的处理单元的附加处理部件 220 处理。部件 220 包括预处理级 230, 其具有亮度空间 238 中的特征检测 232、色度空间 242 中的肤色检测 234 和计分 236。在预处理之后, 编码器 240 对视频数据进行编码, 用于在通信网络上发送。这些部件 220 中的每一个都在随后更具体地描述。

[0028] 给定对图 1 至 2 的多媒体设备 100 的理解, 现在讨论在此公开的用于当可用带宽有限时压缩视频的技术。如前面所指出的, 多媒体设备 100 可以是移动电话或者其它的无线便携式设备。当用于视频会议时, 会议视频中期望的更多细节需要更多的比特经通信网络发送。所有这些最终都是受用于视频会议的通信网络的可用带宽支配的。本公开的技术既使用皮肤检测又使用特征检测来确定要给予视频数据的哪些部分更多或更少的压缩。以这种方式, 可用于设备 100 (例如, 移动电话) 的带宽可以更好地用于把有用的视频数据发送到接收方。

[0029] 参考图 3, 以流程图的形式示出了为利用多媒体设备 100(图 1 至 2) 进行的视频会议压缩视频数据的处理 300。首先, 多媒体设备 100 的处理单元 120 从图像传感器 142 获得视频数据(块 302)。处理单元 120 对视频数据执行肤色检测(块 304) 及特征检测(块 306)。应当指出, 在块 304 和块 306 中所描述的处理步骤可以按任何次序或者甚至同时发生。

[0030] 由肤色和特征检测处理所使用的像素信息依赖于视频数据的色彩空间。在一种布置中, 特征检测在亮度(Y) 空间中进行, 而肤色检测在色度空间(Cb/Cr) 中进行。在另一种实施方式中, 通过简单地直接比较像素颜色值与肤色值, 肤色检测可以直接在 RGB 空间中进行。人与人的肤色或肉色可以不同而且可能基于照明情况而不同。因此, 肤色检测根据需要可以是可配置的或者可调整的。从根本上讲, 处理单元 120 可以利用肤色检测确定视频数据的哪些区域具有属于肤色或肉色范围内的像素值。与此同时、之前或者随后, 处理单元 120 可以确定视频数据的哪些区域具有特征。这些特征可以包括眼睛、鼻子、嘴巴、耳朵、眉毛、手, 等等。

[0031] 知道了这两种类型的区域的定位之后, 处理单元 120 就可以给构成视频数据的各个宏块中的像素的皮肤与特征结果加权或“计分”(块 308)。例如, 肤色区域中的像素与不在这种区域中的其它像素相比将被赋予更高的权重或者重要性。类似地, 特征区域中的像素与不在这种区域中的其它像素相比将被同样地赋予更高的权重或者重要性。最后, 在一个特征中而且在一个肤色区域中或者其附近的像素可以被赋予最高的权重或者重要性。

[0032] 给皮肤与特征结果加权的行可以在视频编解码器 140 的编码器中发生, 而且可以作为把其信息(即, “压缩细节”) 发送到实际编码器的预处理步骤来操作。继而, 视频编解码器 140 的编码器又把“压缩细节”作为编码视频帧不同部分并且把压缩细节转换成它将压缩每一部分的实际比率的指示或者指令。以这种方式, 处理单元 120 中视频编解码器 140 的编码器可以决定如何最好地压缩视频数据以保留细节。

[0033] 根据预处理, 处理单元的视频编解码器 140 的编码器随后基于像素的加权结果执行对区域的压缩(块 310)。于是, 具有肤色的区域将用更多信息(即, 更低压缩) 来编码, 使得这种区域当在远端视频会议设备处被解码时可以更好地被识别。而且, 具有特征的区域将同样用更多信息(即, 更低压缩) 来编码。具有特征并且在肤色区域的阈值距离内的区域将用最多的信息(即, 最低压缩) 来编码。缺乏肤色和特征的其它区域将用较少的信息(即, 更多压缩) 来编码。一旦完成, 处理单元 120 就利用适当的协议和通信电路 114 把编码后

的视频数据在通信网络上发送(块 312)。

[0034] 在探究所涉及的处理的进一步细节之前,首先讨论视频数据如何被处理部件 220 处理。视频数据包括一序列帧 400——图 4 中示出了其中一个。帧 400 被分成像素 420 的块 410。这些块 410 称为宏块而且用于处理帧 400 的视频压缩。通常,宏块 410 的像素宽度与高度是基于四的倍数,但宏块 410 的实际大小依赖于用于编码数据的编解码器。

[0035] 在图 4 中示出的本例中,视频帧 400 是 16 个宏块 410 宽和 16 个宏块 410 高,而且每个宏块 410 是 16x 16 个像素 420。依赖于所采用的二次采样方案,每个像素 420 又可以具有在 YCbCr 色彩空间中的像素信息。这使得视频帧 400 具有 256x 256 像素的像素分辨率。当然,本公开的技术可以应用到任何其它分辨率、宏块大小或者其它布置。例如, HVGA 即“半尺寸视频图形阵列”可以是宽度为二十个 16x 16 的像素宏块,而高度为三十个 16x 16 的像素宏块,具有 480x 320 像素的整体分辨率。

[0036] 如可以看到的,有些宏块(例如,430)包含对象脸部的像素信息并因此具有皮肤或脸部色调。总的来说,如在此所使用的,皮肤或脸部色调可以指与一个人的皮肤或脸关联的任何颜色值。由于人一般是视频会议的最重要的对象,因此所公开的技术集中在肤色或肉色。

[0037] 如还可以看到的,有些宏块(例如,410)不包含对象脸部的像素信息。这些可能或者可能不具有肤色。例如,背景中的宏块 440 可能包含肤色,而其它宏块 450 可能不包含。还有其它宏块(例如,460)可能不包含肤色,尽管它们位于对象的脸中。例如,具有对象眼睛的区域、鼻子下方的区域等的宏块 460 可能在阴影中而且在捕捉到的视频数据中不反映肤色。为了确定如何压缩具有移动/变化对象的一系列帧中复杂视频帧中的各个宏块 410,在此所公开的视频压缩处理的一种实施方式既使用肤色检测又使用特征检测。

[0038] 转向图 5 至 6,同时参考图 2 的部件来描述用于视频压缩的具体处理 500。在 ISP (200) 捕捉到视频数据并且把它转换到 YCbCr 色彩空间等中之后,预处理级(230)的肤色检测处理(234)对一帧视频数据执行肤色检测(块 510)。这种检测可以按本领域中已知的多种方式执行。优选的肤色检测技术的特定细节在于 2009 年 6 月 5 日提交的共同拥有且共同未决的美国申请第 12/479,651 号中公开,该申请的全部内容通过引用并入于此。

[0039] 简单地说,肤色检测识别视频帧中具有肤色的区域。为此,检测处理确定帧中具有属于 CbCr 色彩空间的预定“肤色多边形”或者“肤色楔形”中的颜色值的区域。图 7 至 8 示出了色度空间图 700 和 800,点 702 的假设阵列代表数字图像的像素。点 702 是相对于蓝色色度 Cb 和红色色度 Cr 来映射的。对应于肤色的色调值可以预先确定,而且这种“肤色”区域的边界参数可以编程到存储器中。在色度图 700 中,多边形 704 定义了肤色区域。另一方面,在色度图 800 中,从图的中心发源的“楔形”802 定义肤色区域。

[0040] 色度空间肤色区域确定的每种方法都有其自己的优点。例如,多边形方法更加准确,但计算上更加昂贵。区域定义多边形 704 的边数越大,实现起来计算就越昂贵。另一方面,定义肤色区域的楔形 802 不太准确,而且,尽管楔形 802 中的像素 804 有可能不是肤色,也将它看作是肤色。然而,楔形 802 可以利用向量处理来执行而且可以非常快速地执行。用于定义肤色区域边界的其它技术也可以使用。

[0041] 对于所定义的肤色区域,肤色检测处理(234)可以基于像素的色度 CbCr 值针对视频帧的每个像素作出肤色决定。为此,帧中采样像素的色度值可以经向量处理来检查,以确



定所采样的值把像素放到了预定肤色区域的里面还是外面。每个宏块的每个像素都可以被检查而且可以把代表其肤色决定的值(例如,“肤色”或“非肤色”的二进制决定)写到存储器(例如,存储器 116)中。可选地,可以只检查一个或多个宏块中的一些像素,来提高处理速度。

[0042] 回到图 5,特征检测处理(232)检测在帧的视频数据中所定义的特征或者特定区域(块 515)。这种检测可以在肤色检测之前、期间或者之后执行,而且也可以按本领域中已知的多种方式来执行。对给定像素是否为“特征”的一部分的确定可以基于任意数量的众所周知的特征或边缘检测算法,例如 Viola Jones 对象检测或者 Sobel 边缘检测。每个宏块的每个像素都可以被检查,并且可以把代表其特征决定的值(例如,“特征”或“非特征”的二进制决定)写到存储器(例如,存储器 116)中。可选地,可以只检查一个或多个宏块中的一些像素,来提高处理速度。

[0043] 简单地说,特征检测处理(232)识别帧中包含特征或边缘的像素。为此,特征检测处理可以使用亮度(Y)空间中的像素值。例如,公共边缘检测算法尝试识别数字图像中图像亮度(Y)值逐像素变化最显著或者图像有中断的点和方向。边缘检测可以设置任何数量的阈值参数,以确定给定的特征(即,边缘)在数字图像中的什么地方开始和结束。如图 9 中所示,数字图像 900 中假设定位的边缘由图像上的散列线 910 表示。注意,特征检测处理也有可能定位背景对象中的一些边缘或特征 920。对于在此所述的方法,重要的是能够区分对视频会议压缩更加重要的特征 910 和对视频会议压缩不太重要的特征 920,以使得包括视频流的比特可以花在最重要的位置。

[0044] 如图 10 中所示,数字图像 1000 中对视频会议更重要的区域,即,图像中作为特征的一部分而且还在肤色区域的阈值距离之内的那些区域,在图像上保持由散列线 910 表示。数字图像 1000 中不太重要的区域,即,图像中不包含特征或者包含不位于肤色区域的阈值距离内的特征的那些区域,在图像上不再由散列线 910 表示。数字图像 1000 中保持由散列线 910 表示的区域对应于数字图像 1000 中将赋予最高压缩得分并将以最低比率被视频编码器压缩的宏块。

[0045] 回到图 5,一旦发生了肤色与特征检测,计分处理(236)就可以既使用检测到的肤色区域又使用检测到的特征来确定定义如何压缩帧的视频数据的得分或权重。首先,选择对象帧的宏块(块 520),并检查所选宏块的像素(块 530)。计分处理(236)使用所存储的用于每个像素的肤色与特征决定来确定所选的像素是只在肤色区域内、只在特征区域内,还是既在特征区域内又在肤色区域的阈值距离内(判决 532/534/536)。依赖于分析,计分处理(236)相应地给像素加权或计分(块 540/542/544/546)。

[0046] 在一个例子中,赋予宏块中的每个像素一个加权值,这是基于像素:a)是在肤色区域内(例如,值为“1”);b)是“特征”的一部分(例如,值为“2”);c)是“特征”的一部分并且在肤色区域的阈值距离(例如,两个像素)内(例如,值为“16”)或者 d)既不在肤色区域内也不是“特征”的一部分(例如,值为“0”)。如以上所提到的,所使用的实际计分系统可以是可配置的而且可以依赖于期望的实现和带宽约束。

[0047] 最后,处理 500 确定是否还有更多像素、宏块和帧(判决 550/552/554)。处理 500 可以处理视频数据的每个帧或者可以只处理视频数据的选定帧,使得不需要处理每个帧来确定压缩变量。换句话说,假定视频会议中的对象没有在改变位置,那么视频数据的一帧可

以由处理 500 来处置,然后多个后续的帧可以使用相同的确定。在给定的条件下,这可能是或者可能不是有用的。同样,不同帧中不同的宏块可以持续地利用处理 500 和用于进入的帧的确定来处置。

[0048] 不管宏块和帧是如何处置的,处理 500 都可以象图 6 中所示出的那样继续进行。计分处理(236)取得要压缩的对象帧(块 560)并取得该帧的一个宏块(块 562)。利用先前确定的只在肤色区域中、只在特征中或者既在特征区域中又在肤色区域的阈值距离内的像素的得分,计分处理(236)为该宏块计算平均像素值得分。

[0049] 在平均该帧的每个或者一些宏块(判决 566)之后,计分处理(236)确定用于宏块的压缩得分(块 570)。为此,计分处理(236)利用用于给定宏块的平均像素值来计算压缩得分。总的来说,压缩得分可以基于预定的数值范围(例如,1 至 10 的数值范围)。在一种实施方式中,具有压缩得分 10 的宏块将被视频编解码器压缩最少,而具有压缩得分 1 的宏块将被视频编解码器压缩最多。

[0050] 压缩得分可以通过把每个宏块的平均像素值线性缩放到期望的数值范围(例如,1 至 10)来确定。例如,给定视频帧中具有最高像素平均值的宏块可以设置成具有最高压缩得分(例如,“10”),而剩余的宏块可以根据某种适当的函数相对于其来缩放。可选地,压缩得分可以简单地设置成用于该宏块的平均像素值。

[0051] 为了节省某些附加的带宽,如果没有宏块达到某个阈值像素平均值,那么就不赋予任何宏块压缩得分“10”。以这种方式,信息将不会仅仅因为它具有当前帧中最好或最高的得分而用最低压缩来编码。这种当前帧中的最佳得分信息可能不是视频会议的参与者的信息,因此高压缩可能不合适或者不必要。例如,照相机可以指向到白色的墙壁,在这种情况下,高度压缩帧中的所有宏块将是合适的和优选的。但是,物体(即,纸张、展品等)可能实际上是视频会议的当前帧中的期望对象,因此这个特征可以可选地实现和/或是可配置的。

[0052] 一旦确定了用于期望宏块的压缩得分,编码器(240)就基于该压缩得分压缩宏块(块 572)并继续到下一个帧(判决 574)。例如,压缩得分可以以视频帧附件的形式传递到编码器。帧中具有较少有用信息的区域比具有更多有用信息(例如,肤色和特征)的区域压缩得更多。因而,所述方法朝着对视频会议有最大利益的区域,即,脸、手而且尤其是脸部特征,优化用于视频帧上宏块的压缩比的分布。编码器 240 可以使用用于压缩视频会议视频的任何合适的标准格式。一些标准包括 H. 263 和 H. 264 视频编解码器标准。

[0053] 在会议过程中,基于其对视频图像的实际重要性,在视频中可能发生一个视频帧的部分被压缩多于或少于其“应该”的情况。例如,这些部分或者“离群值宏块”可能因为属于特定宏块的对象脸的一部分位于阴影中使得该宏块被压缩多于周围宏块而出现,因为所述方法不能确定阴影区域是肤色和/或特征的一部分。这将导致给定的宏块是漏判(False Negative)。相反的情况可以称为误判(False Positive),即,因为某种照明条件、边缘效应或者二者结合使得宏块在计分处理看起来对视频帧比实际情形更重要而使该宏块比其周围宏块压缩较少的情况。例如,图像背景中的树木或者植物可能具有带许多边缘的区域和/或离肤色足够近使得它们得到非常高的压缩得分,尽管这种确定或者是不准确的或者对整体的压缩方案无益。

[0054] 为了处理这些情况,在已经赋予每个宏块初始压缩得分之后可以二次处置处理后

的帧的宏块。如图 11 中所示,这个继续的处理 1100 取得已经处理过的帧(块 1102) 并且选择宏块(块 1104)。然后,处理 1100 比较选定的宏块的压缩得分与其邻近宏块的得分(块 1106)。这么做是为了找出离群值,例如被高压缩宏块包围的低压缩宏块(即,误判) 或者被全部低压缩宏块包围的高压缩宏块(即,漏判)。

[0055] 例如,图 12 示出了一个对象的视频帧 1200,该帧具有与位于深阴影中的对象眼睛周围隔离的一组宏块 1202。中央宏块 1210A-B 代表漏判。与其周围的邻居(即,八个相邻的宏块) 的得分(例如,“5”、“7”& “10”) 相比,这些宏块 1210A-B 具有低得分(例如,“3”)。尽管这些宏块 1210A-B 包括代表对象眼睛的像素,即,用于视频会议的视频帧的重要区域,但是低得分的宏块 1210A-B 将比其邻居压缩得多。可以假定,得分中的差异是由于深阴影、色差、不可靠的边缘检测等导致的。不管是出于什么原因,把对象的眼睛比周围的眉毛和脸颊压缩更多将是不太理想的。为此,使用所公开处理的二次进行来检测这种离群值并校正它们。

[0056] 误判可能由于多种原因而在帧 1200 中发生。例如,帧 1200 可能包含背景中的森林或者树木,而且可能有或者可能没有来自树枝与树叶细节的许多特征和边缘。此外,颜色可能足够接近肤色,使得给定宏块 1220 的得分可以导致相对低的压缩比。例如,给定宏块 1220 可以类似于真正包含对象脸部部分的另一个宏块 1230 而被计分。如果使用这种压缩得分,那么将赋予宏块 1220 比所需更多的重要性。通过查看周围宏块的得分,可以智能地调整给定宏块的得分。

[0057] 回到图 11,继续处理 1100 比较选定宏块的压缩得分与其邻近宏块的得分(块 1106)。选择哪些及多少邻近的宏块与选定宏块进行比较可以依赖于环境、得分之间的差异,而且可以是可配置的。如果大部分、全部或者某个水平的邻居得分高于选定宏块的得分,处理 1100 就确定这个宏块是漏判离群值(判决 1110),而且处理 1100 调整该选定宏块的得分,以反映其邻居的得分(块 1114)。可选地,如果大部分、全部或者某个水平的邻居得分低于选定宏块的得分,处理 1100 就确定这个宏块是误判离群值(判决 1112),而且处理 1100 调整该选定宏块的得分,以反映其邻居的得分(块 1114)。最终,在进行到下一个或某个后续的视频帧(块 1102) 之前,处理 1100 选择每个宏块(或者其某个选定集合) (判决 1116)。

[0058] 在一种实施方式中,可以对期望的宏块执行误判 / 漏判离群值得分校正处理。这个处理可以定义一个用于漏判宏块的阈值,例如“5”。然后,该处理可以检查选定宏块的每个邻近宏块的压缩得分。如果该选定宏块的邻近宏块的平均压缩得分等于或者高于漏判阈值,例如“5”,而且该选定宏块的得分小于漏判阈值,处理就把选定宏块的压缩得分设置成该选定宏块的邻近宏块的平均压缩得分。同样,校正处理可以定义另一个用于误判宏块的阈值,例如“2”。然后,该处理可以检查选定宏块的每个邻近宏块的压缩得分。如果用于该选定宏块的邻近宏块的平均压缩得分等于或者小于误判阈值,例如“2”,而且该选定宏块的得分高于误判阈值,处理就把选定宏块的压缩得分设置成该选定宏块的邻近宏块的平均压缩得分。如所期望的,用作误判和漏判阈值的值当然可以调整,用于对误判 / 漏判离群值宏块更积极或更不积极的处理。选择例如“2”的误判阈值以确保图像背景中的噪声不干扰误判的去除。另一方面,如果宏块具有至少例如“5”的较高压缩得分,那么就可以假定该宏块拥有处于皮肤区域内的足量肤色像素。

[0059] 如以上所提到的,在调整选定宏块得分时(块 1114),处理 1100 可以利用邻近宏块

得分的平均值并指定选定宏块具有该平均得分。可选地,如果期望的话,处理可以使用某种其它的预设得分。邻近宏块的得分还可以基于与选定宏块的接近程度和基于得分之间的差异来加权。也可以执行某种其它的确定。最终,这个二次进行的处理 1100 确定选定宏块是否是应当校正其压缩得分的漏判或误判离群值。

[0060] 在此所述的智能肤色与脸部特征识别视频会议压缩方法还可以具有在非视频会议应用中的一些应用,例如监控摄像头。监控系统存储很长时间内的非常大量的数据,因而,可能期望以比墙壁或其它背景物体低的比率压缩人的脸(例如,银行的监控摄像头)。

[0061] 以上对优选和其它实施方式的描述不是要限制或约束本申请人设想的创新概念的范围或适用性。以公开在此所包含的创新概念为交换,本申请人期望由所附权利要求提供的所有专利权。因此,所附权利要求旨在最大程度地包括对属于以下权利要求或者其等同物范围之内全部修改与变更。

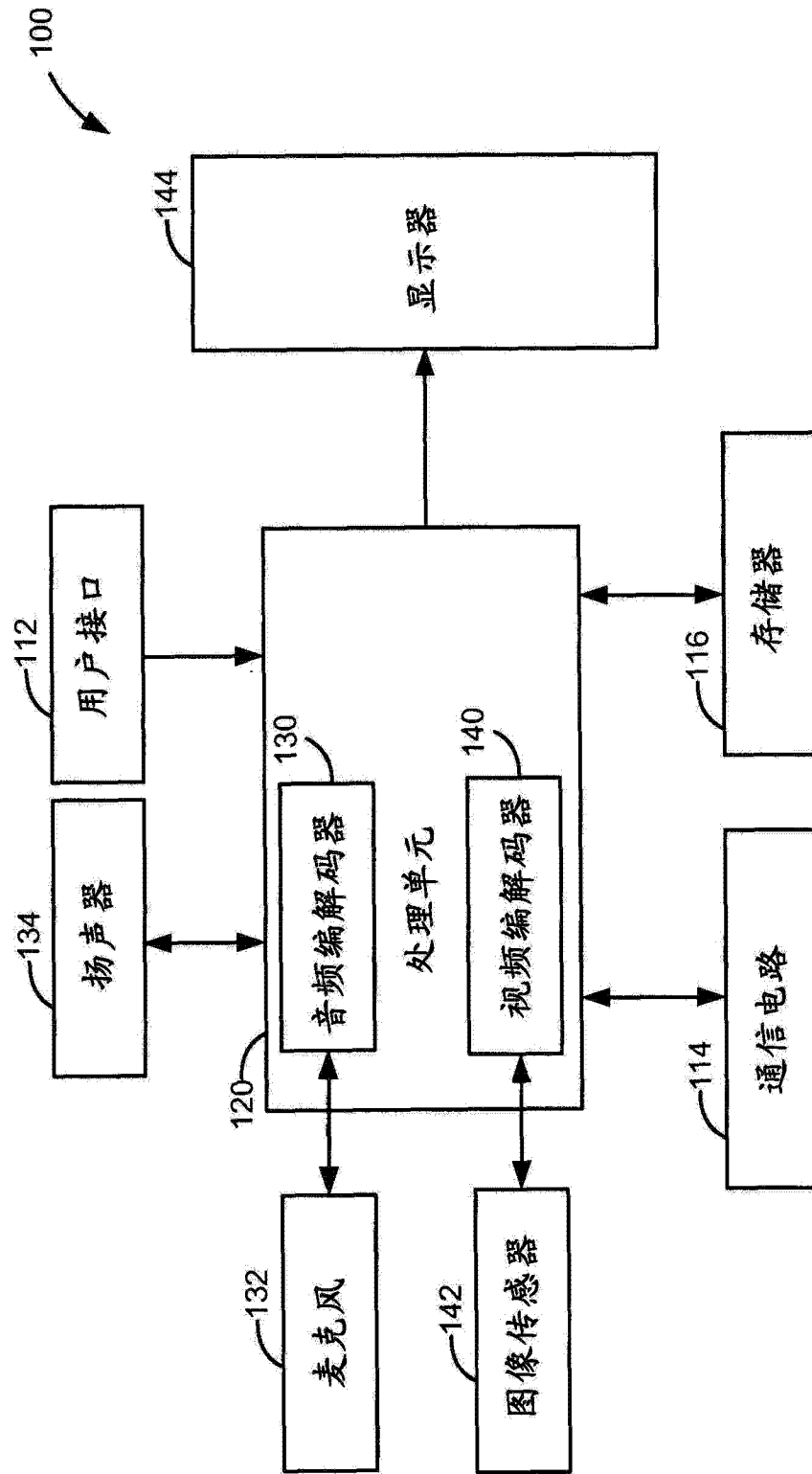


图 1

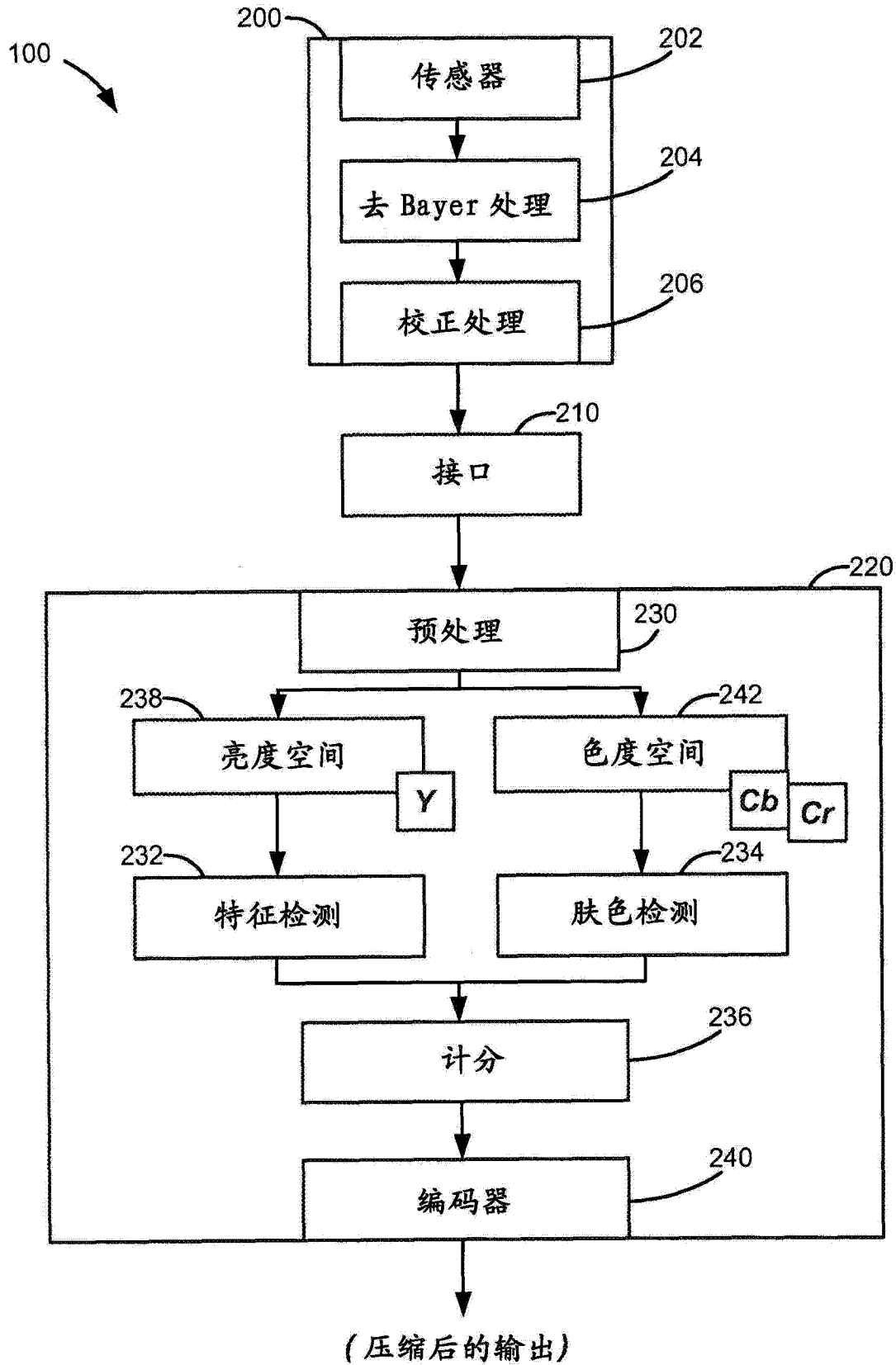


图 2

300

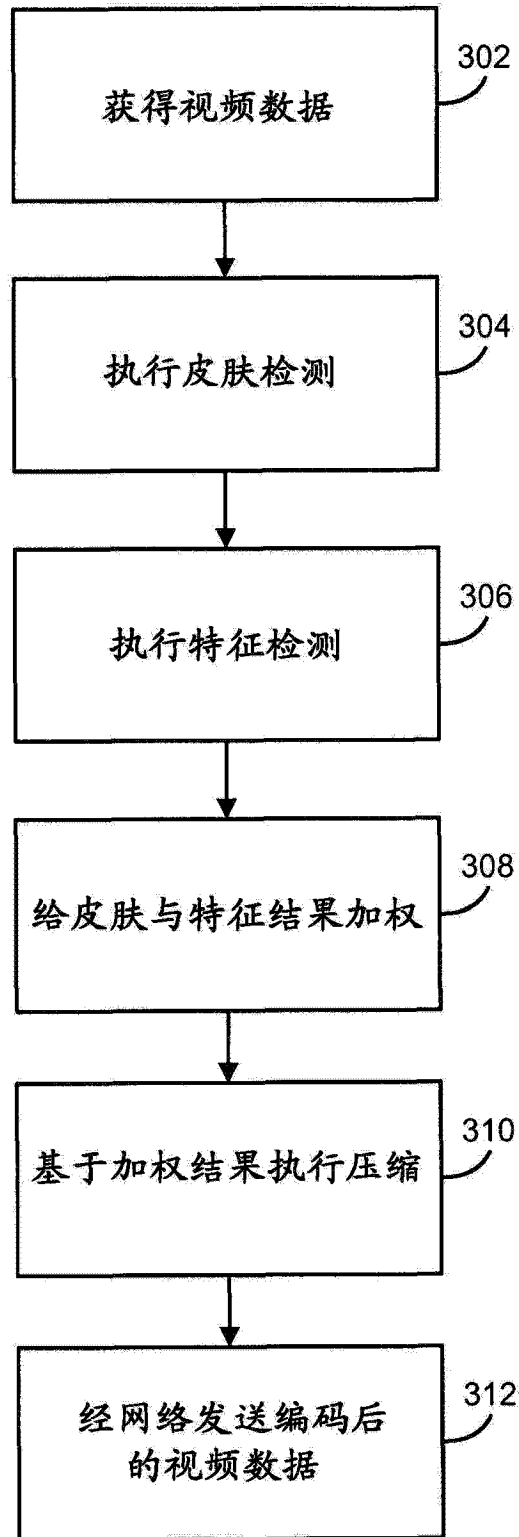


图 3

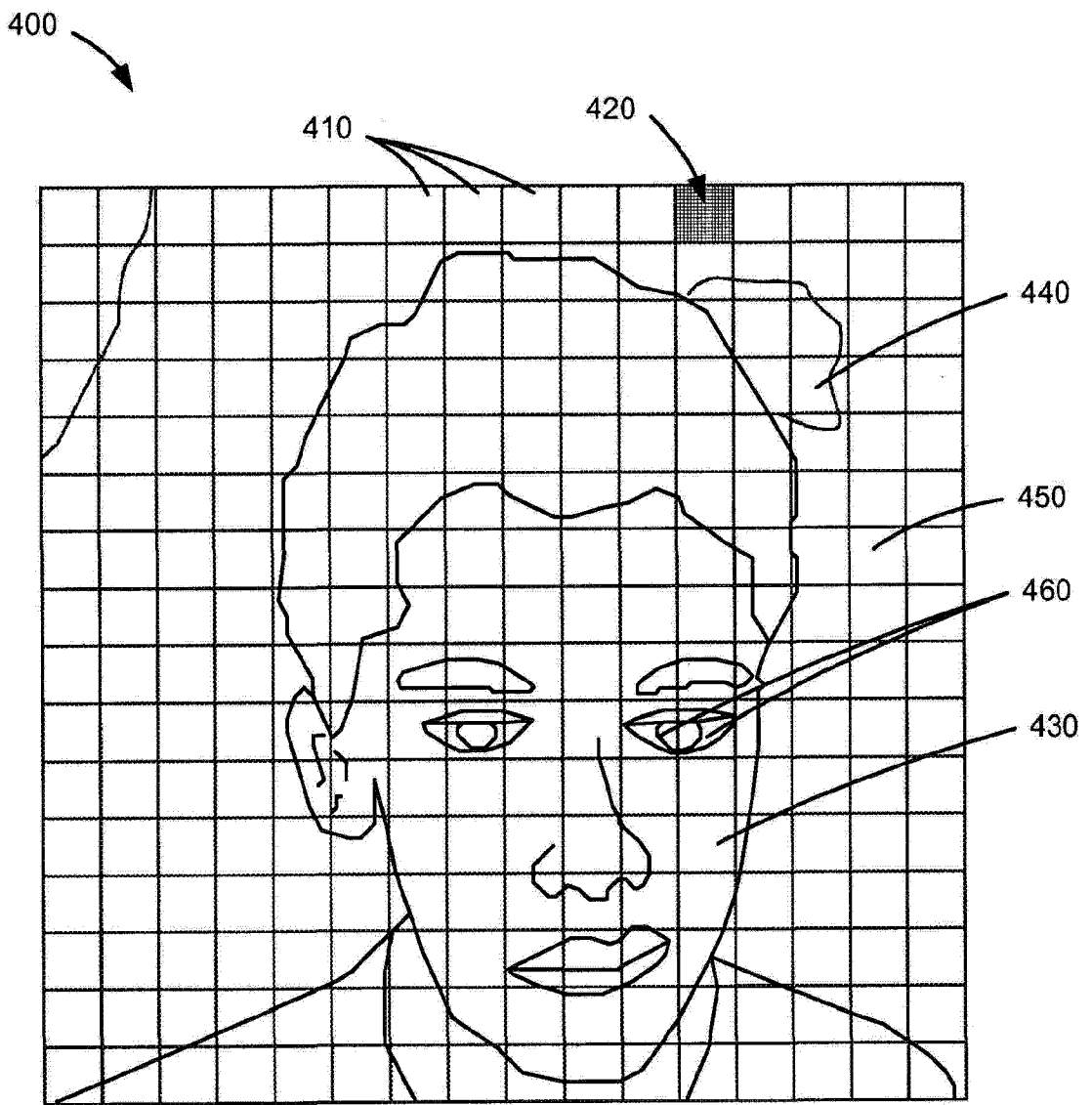


图 4



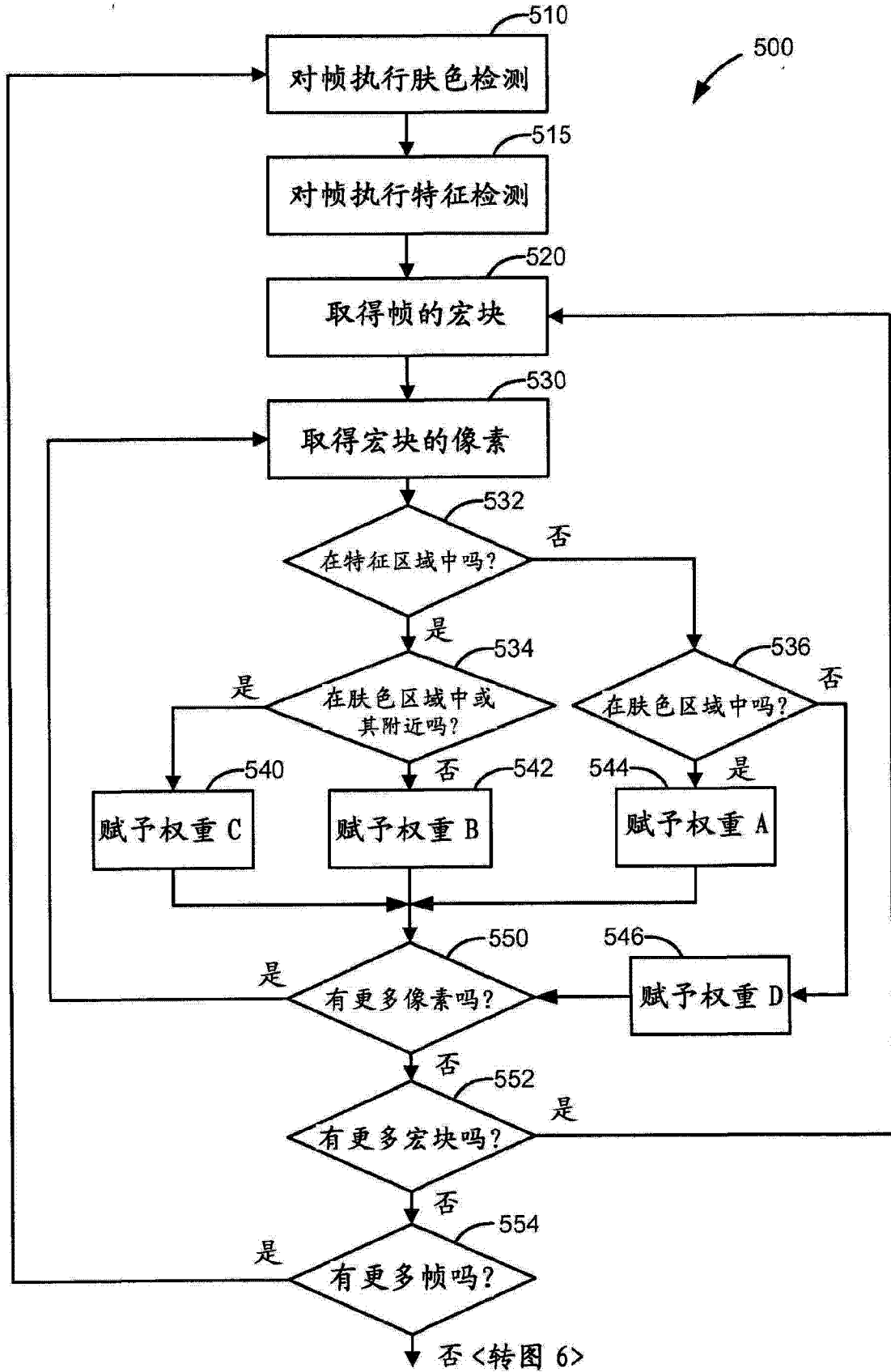


图 5

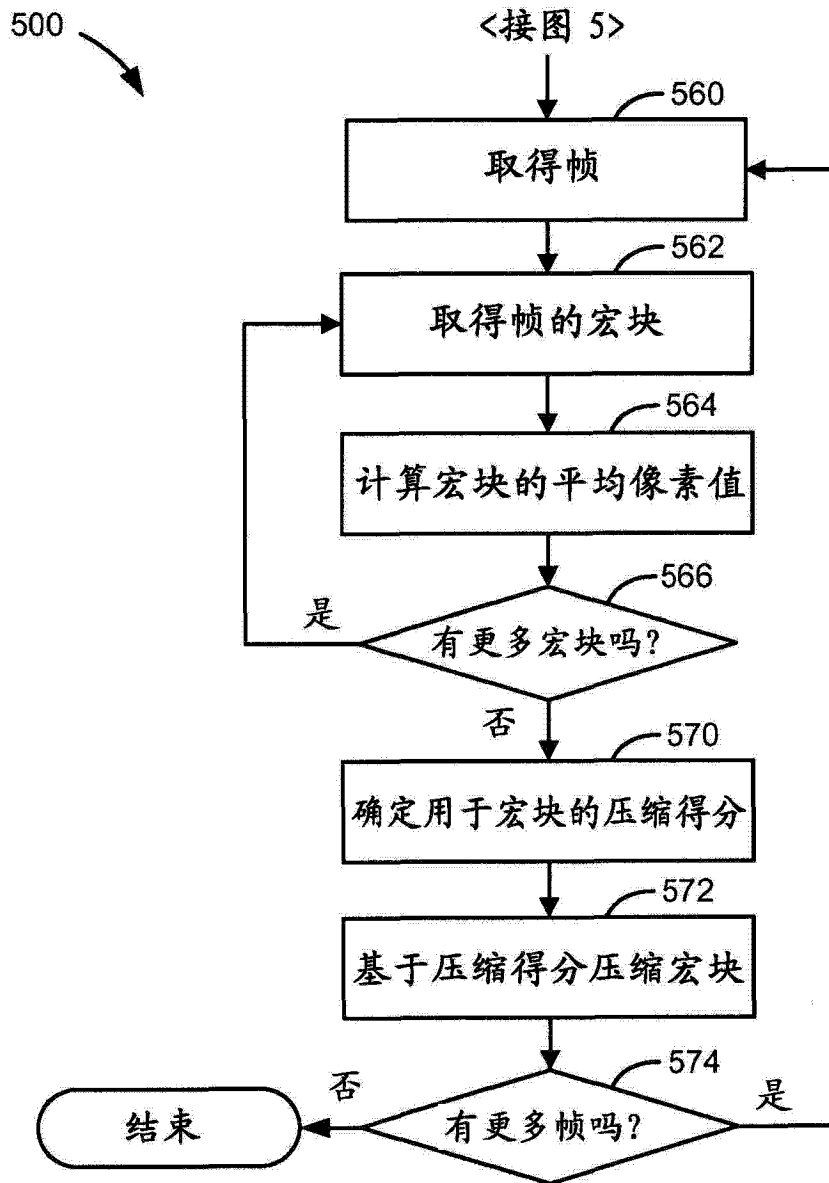


图 6

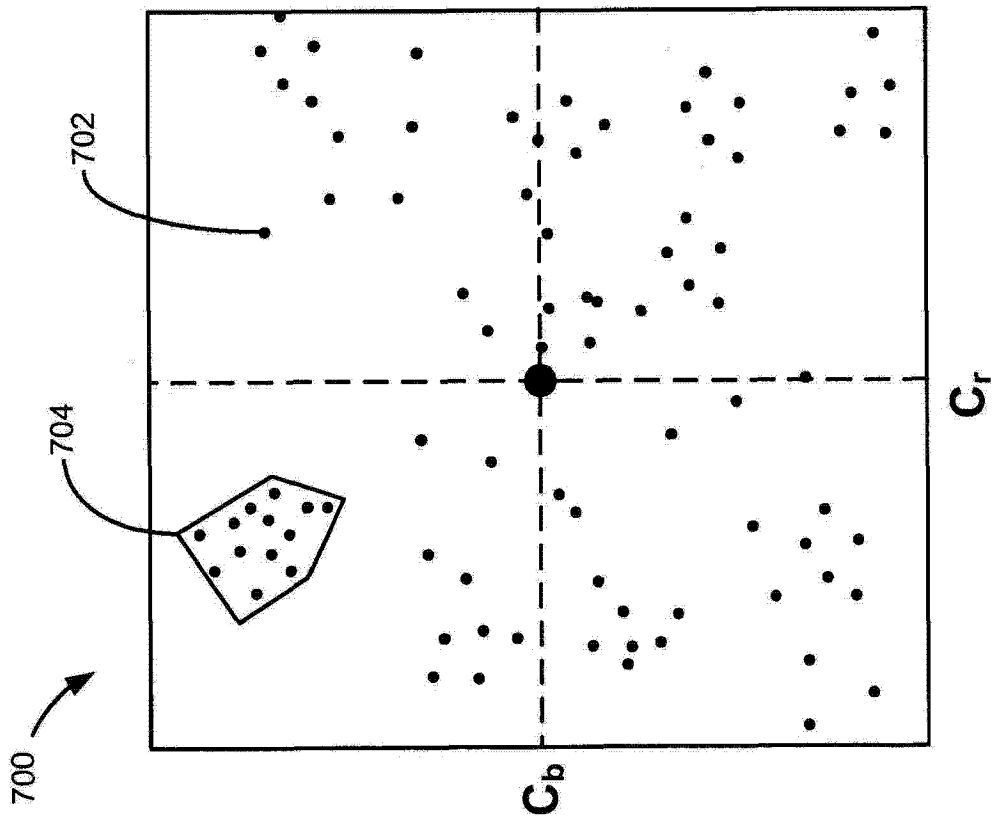


图 7

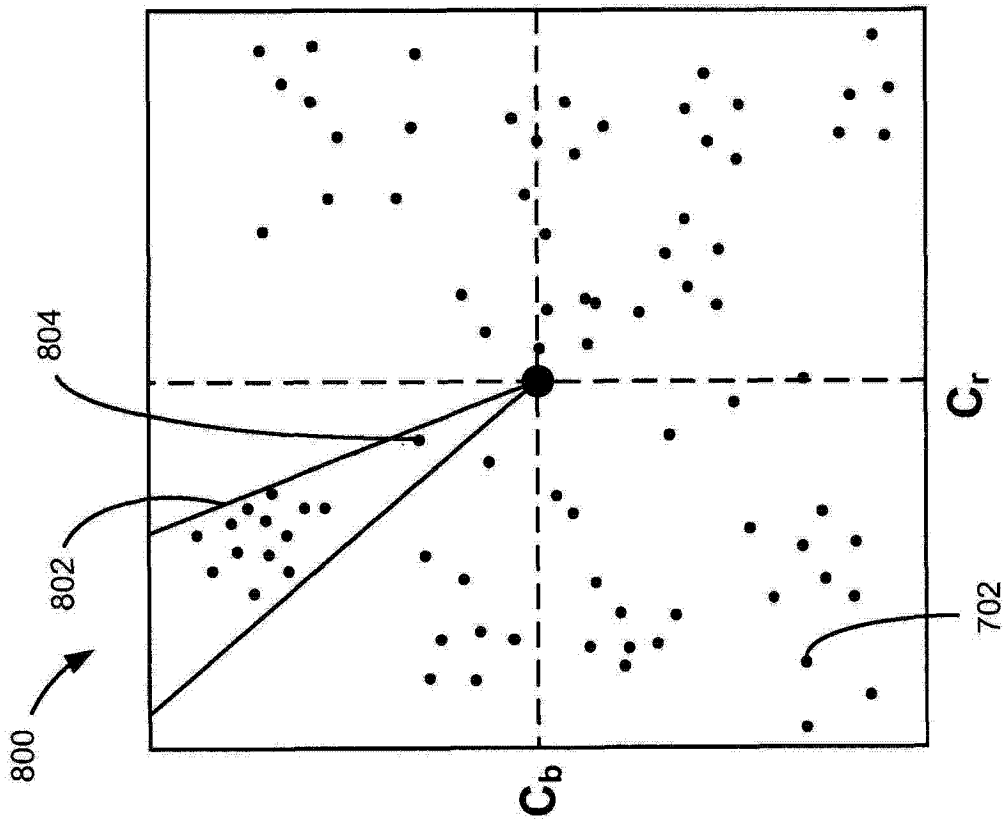


图 8

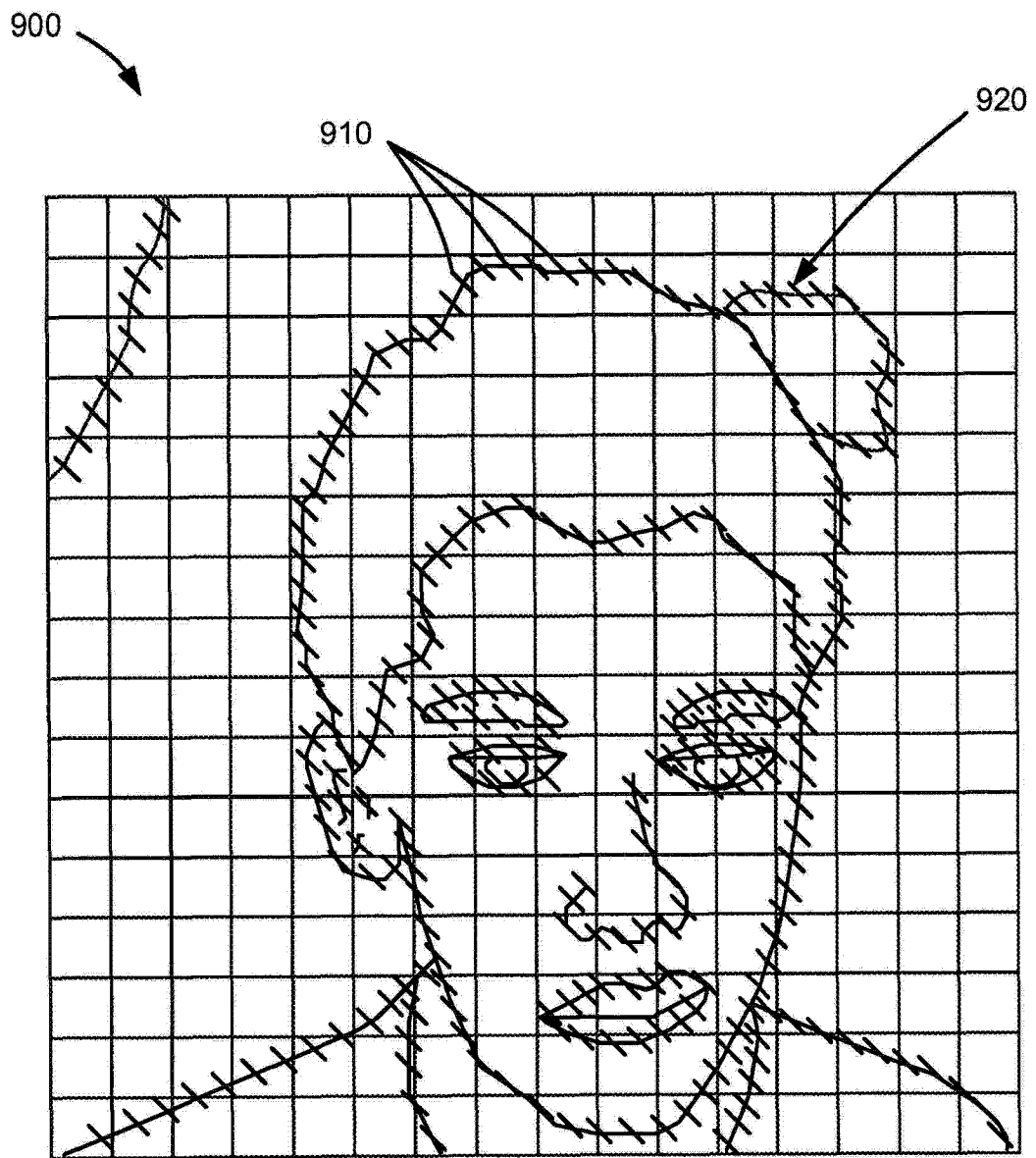


图 9

1000

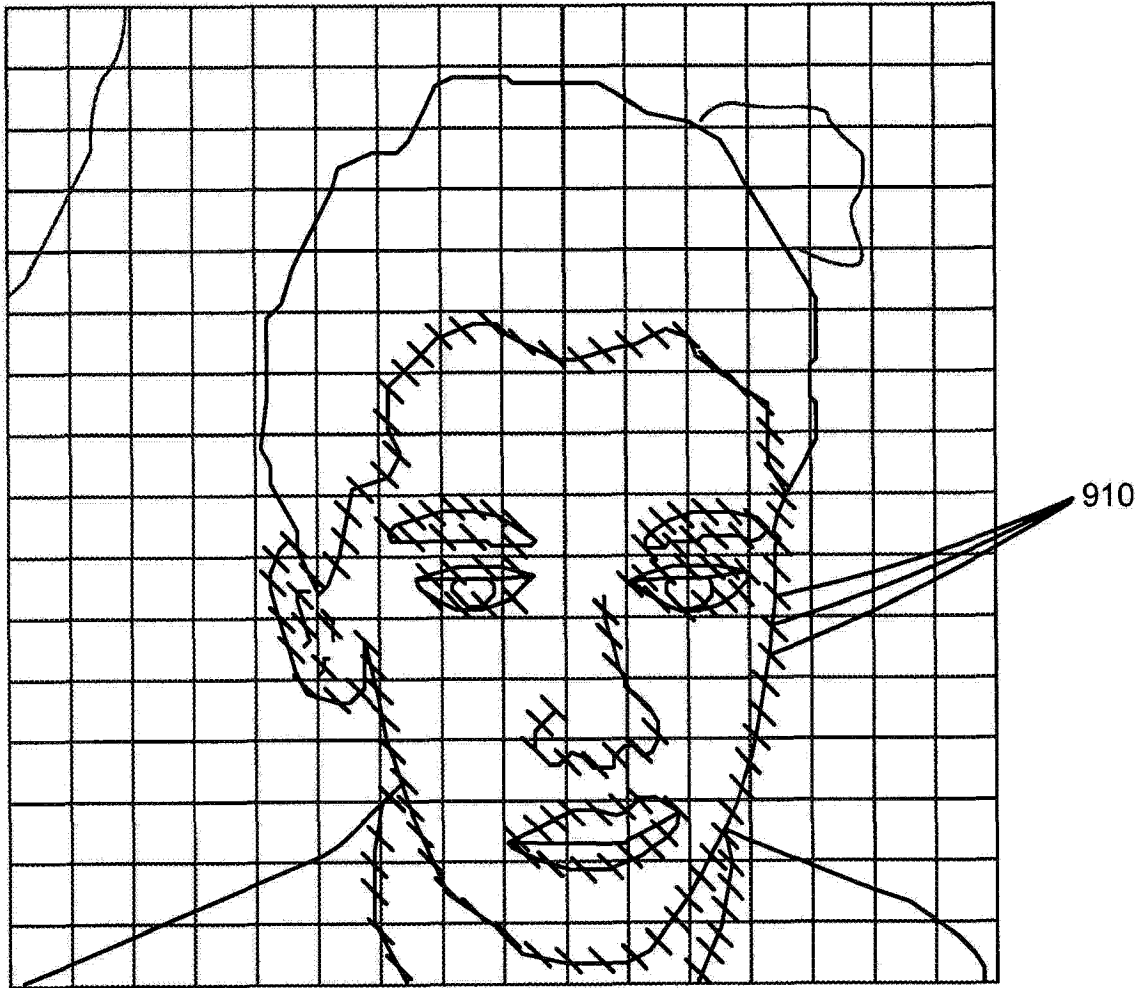


图 10

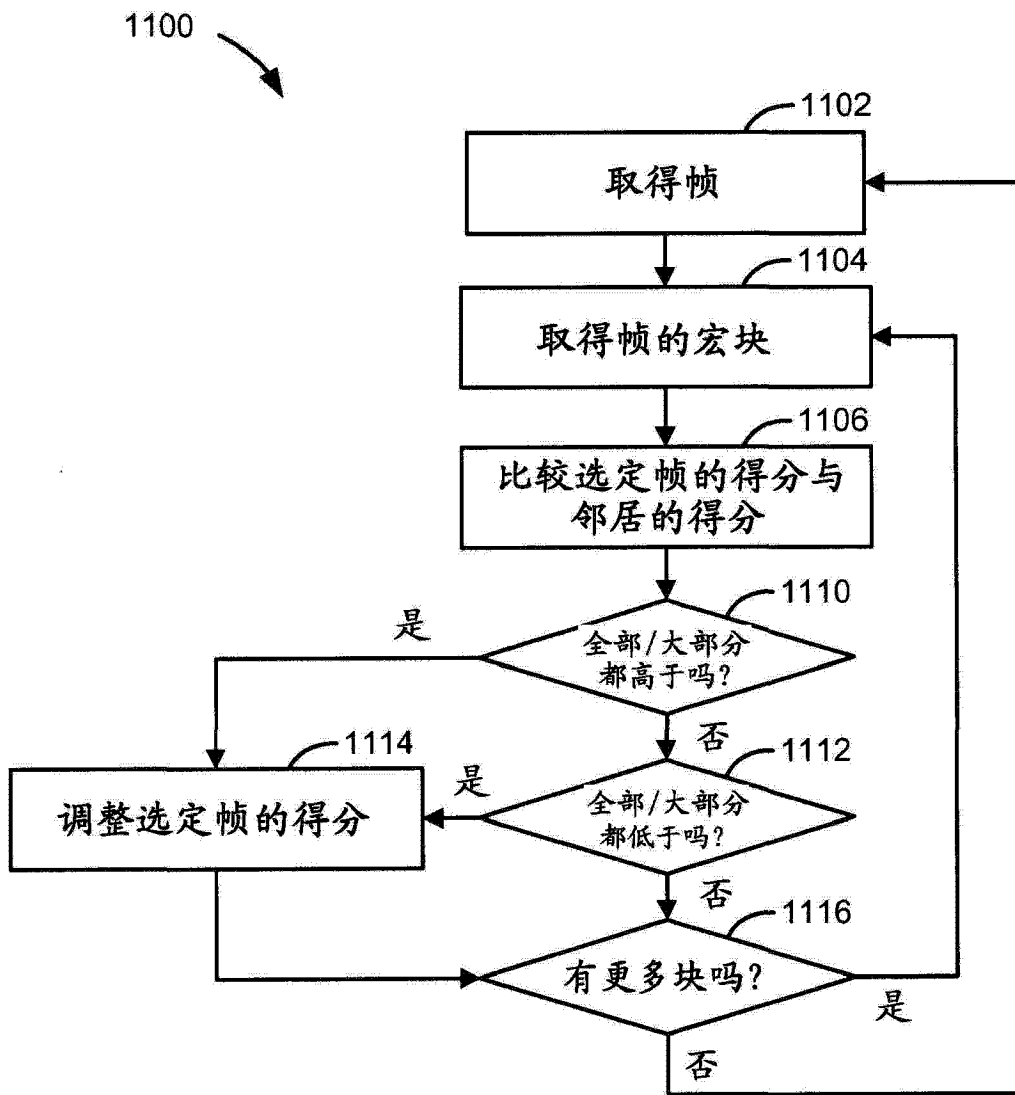


图 11

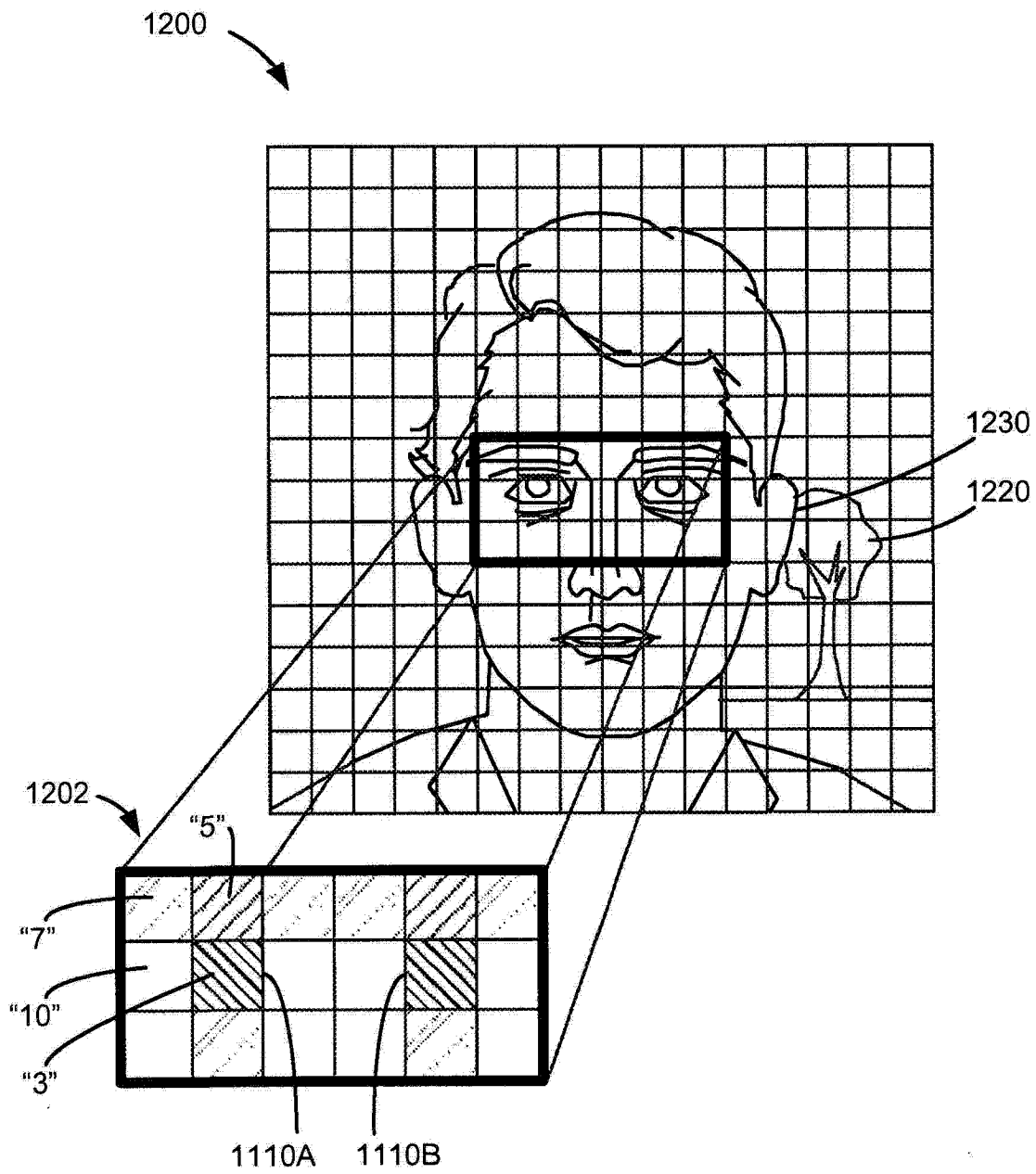


图 12