



(12) 发明专利申请

(10) 申请公布号 CN 103098076 A

(43) 申请公布日 2013. 05. 08

(21) 申请号 201180043241. 9

(51) Int. Cl.

(22) 申请日 2011. 07. 26

G06K 9/00 (2006. 01)

(30) 优先权数据

12/885, 072 2010. 09. 17 US

(85) PCT申请进入国家阶段日

2013. 03. 08

(86) PCT申请的申请数据

PCT/US2011/045287 2011. 07. 26

(87) PCT申请的公布数据

W02012/036790 EN 2012. 03. 22

(71) 申请人 索尼公司

地址 日本东京

(72) 发明人 黄光满 刘明昌 余良吟

(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 张荣海

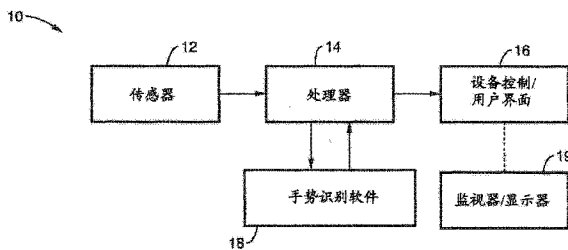
权利要求书3页 说明书9页 附图10页

(54) 发明名称

电视机控制用手势识别系统

(57) 摘要

一种利用与运动信息结合的基于肤色的方法,实现实时分割的手势识别系统。利用 Kalman 滤波器跟踪手部的质心。计算提取的手部掩膜的手掌中心,手掌底部,以及从手掌中心到轮廓的最大距离。随后比较计算的距离和阈值,以确定当前姿势是“张开”还是“握紧”。在优选实施例中,“张开”姿势和“握紧”姿势之间的转变确定当前手势是处于“选择”状态还是“抓取”状态。



1. 一种用于遥控设备的手势识别系统,包括:
  - (a) 在所述设备附近的位置处,捕捉用户的手部的视频数据的传感器;
  - (b) 处理捕捉的所述手部的视频数据的处理器;
  - (c) 可在所述处理器上运行,以执行多个步骤的程序,所述多个步骤包括:
    - 根据肤色,分割捕捉的视频中的每个手部图像;
    - 提取分割的手部图像的一个或多个参数;
    - 跟踪手部图像中的所述一个或多个参数;
    - 根据所述一个或多个参数,把手势分类成“张开”或“握紧”;和
    - 根据所述手势的识别和对所述一个或多个参数的位置的跟踪,操纵所述设备。
2. 按照权利要求 1 所述的系统,其中分割每个手部图像包括:
  - 把与所述手部相关的前景区域和围绕所述前景区域的背景区域分开;
  - 训练背景区域和前景区域的 3 分量混合模型;和
  - 根据运动和肤色中的一个或多个,对图像像素分类。
3. 按照权利要求 1 所述的系统,其中只利用 Cb 和 Cr 颜色通道分割图像。
4. 按照权利要求 1 所述的系统,其中提取一个或多个参数包括:
  - 根据提取的手部的轮廓和骨架,获得用户的手部的手掌中心位置。
5. 按照权利要求 4 所述的系统,其中跟踪一个或多个参数包括利用 Kalman 滤波器,跟踪手掌中心位置。
6. 按照权利要求 4 所述的系统,其中提取一个或多个参数还包括:
  - 获得手掌底部位置;和
  - 测量手掌中心位置与提取的图像轮廓上的最远轮廓点之间的距离;
  - 所述最远轮廓点在通过手掌底部位置和手掌中心位置的矢量的预定角度范围内。
7. 按照权利要求 6 所述的系统,其中把手势分类成“张开”或“握紧”包括:
  - 比较到最远轮廓点的距离和阈值;
  - 如果到最远轮廓点的距离超过所述阈值,那么把图像分类成“张开”;以及
  - 如果到最远轮廓点的距离低于所述阈值,那么把图像分类成“握紧”。
8. 按照权利要求 1 所述的系统,
  - 其中所述设备包含与显示器耦接的用户界面,所述用户界面是在所述显示器上输出的;以及
  - 其中操纵所述设备包括根据所述手势的识别和所述一个或多个参数的跟踪,向用户界面发送命令。
9. 按照权利要求 8 所述的系统,其中“张开”手势和“握紧”手势之间的转变被用于确定“选择”或“抓取”命令是否被发送给用户界面。
10. 按照权利要求 9 所述的系统,其中张开手势的跟踪被用于确定光标在用户界面上的位置。
11. 一种用于遥控设备的手势识别系统,所述设备具有在显示器上显现的用户界面,所述手势识别系统包括:
  - (a) 在所述设备附近的位置,捕捉用户的手部的视频数据的传感器;
  - (b) 处理捕捉的所述手部的视频数据的处理器;

(c) 可在所述处理器上运行以执行多个步骤的程序,所述多个步骤包括:  
根据肤色,分割捕捉的视频中的每个手部图像;  
提取分割的手部图像的一个或多个参数;  
跟踪手部图像中的所述一个或多个参数;  
根据所述一个或多个参数,把手势分类成“张开”或“握紧”;和  
根据所述手势的识别和所述一个或多个参数的跟踪,操纵所述设备;  
其中操纵所述设备包括根据所述手势的识别和对所述一个或多个参数的位置的跟踪,  
向用户界面发送命令。

12. 按照权利要求 11 所述的系统,其中分割每个手部图像包括:  
把与所述手部相关的前景区域和围绕所述前景区域的背景区域分开;  
训练背景区域和前景区域的只包含 Cb 和 Cr 颜色通道的 3 分量混合模型;和  
根据运动和肤色中的一个或多个,对图像像素分类。

13. 按照权利要求 11 所述的系统,其中提取一个或多个参数包括:  
根据提取的手部的轮廓和骨架,获得用户的手部的手掌中心位置;  
根据手掌中心位置,获得手部的手掌底部位置;和  
测量手掌中心位置与提取的图像轮廓上的最远轮廓点之间的距离;  
所述最远轮廓点在通过手掌底部位置和手掌中心位置的矢量的一定角度范围内。

14. 按照权利要求 13 所述的系统,其中跟踪所述一个或多个参数包括利用 Kalman 滤波器跟踪手掌中心位置。

15. 按照权利要求 13 中所述的系统,其中把手势分类成“张开”或“握紧”包括:  
比较到最远轮廓点的距离和阈值;  
如果到最远轮廓点的距离超过所述阈值,那么把图像分类成“张开”;以及  
如果到最远轮廓点的距离低于所述阈值,那么把图像分类成“握紧”。

16. 按照权利要求 15 所述的系统,其中“张开”手势和“握紧”手势之间的转变被用于确定“选择”或“抓取”命令是否被发送给用户界面。

17. 按照权利要求 16 所述的系统,其中张开手势的跟踪被用于确定光标在用户界面上的位置。

18. 一种利用手势遥控设备的方法,所述设备具有在显示器上显现的用户界面,所述方法包括:

用在所述设备附近位置的传感器,捕捉用户的手部的视频;  
根据肤色,分割捕捉的视频中的每个手部图像;  
提取分割的手部图像的一个或多个参数;  
跟踪手部图像中的所述一个或多个参数中的至少一个参数;  
根据所述一个或多个参数,把手势分类成“张开”或“握紧”;  
根据所述手势的识别和所述一个或多个参数的跟踪,操纵所述设备;  
根据所述手势的识别和所述一个或多个参数的跟踪,向用户界面发送命令;和  
在显示器上输出发送的命令,以便操纵所述设备。

19. 按照权利要求 18 所述的方法,  
其中提取一个或多个参数包括:

根据提取的手部的轮廓和骨架,获得用户的手部的手掌中心位置;  
根据手掌中心位置,获得手部的手掌底部位置;和  
测量手掌中心位置与提取的图像轮廓上的最远轮廓点之间的距离;  
所述最远轮廓点在通过手掌底部位置和手掌中心位置的矢量的一定角度范围内,其中把手势分类成“张开”或“握紧”包括:  
比较到最远轮廓点的距离和阈值;  
如果到最远轮廓点的距离超过所述阈值,那么把图像分类成“张开”;以及  
如果到最远轮廓点的距离低于所述阈值,那么把图像分类成“握紧”。

20. 按照权利要求 19 所述的方法,其中“张开”手势和“握紧”手势之间的转变被用于确定“选择”或“抓取”命令是否被发送给用户界面;以及  
其中对张开手势的手掌中心的跟踪被用于确定光标在用户界面上的位置。

## 电视机控制用手势识别系统

[0001] 相关申请的交叉引用

[0002] 本申请要求 2010 年 9 月 17 日提交的美国专利申请 12/885072 的优先权,通过引用将其全部内容并入于此。

[0003] 受版权保护的素材公告

[0004] 本专利文献中的一部分素材受美国和其他国家版权法的版权保护。版权所有人不反对任何人影印再现专利文献或专利公开,因为专利文献或专利公开出现在可公开获得的美国专利商标局文件或档案中,不过在别的方面却保留所有版权。版权所有者从而不放弃其使本专利文献保密的任何权利,包括(但不限于)其按照 37C. F. R. § 1. 14 的权利。

### 技术领域

[0005] 本发明涉及手势识别,更具体地说,涉及遥控用手势识别。

### 背景技术

[0006] 手势识别是一种新兴技术,它能够提供更直接的人机交互方法。手势识别与空间和时间信息有关,可用于为未来的应用替代传统的控制设备。已尝试了手势识别系统,例如,美国专利 No. 5, 534, 917。不过,这些系统通常涉及计算密集的复杂分析和/或需要昂贵的硬件。

[0007] 因而,本发明的目的是一种能够从典型的 HD 照相机获得输入,和利用一组简单的手势使计算力降至最低的手势识别系统。在下面的说明中,将满足这些目标中的至少一些目标。

### 发明内容

[0008] 本发明包含处理作为输入的运动手部的视频,并输出每个图像中的当前手势状态和手部位置的系统和方法。在每个图像中,从背景中分割手部区域,并根据一个或多个参数,将其分类为“张开的手部”或“握紧的手部”。系统结合肤色模型和运动中的一个或多个,以获得实时手部分割性能。典型的应用包括(但不限于)用于基本电视机控制,电视机上的照片和 web 浏览,和计算机的手势识别。

[0009] 本发明在仅仅利用单个 HD 照相机的同时,以较低的计算复杂性,实现用于电视机控制的实时并且用户友好的手势识别系统。

[0010] 系统利用小型传感器进行手势识别,可利用简单手势,在 3 米以上的距离控制电视机、计算机或其它设备。组合肤色模型和运动,以便从背景中分离手部区域

[0011] 操作中,提取几个参数,比如手部大小,手掌中心和手部底部。这些参数被用于把当前手势分成“张开的手部”或“握紧的手部”。系统随后利用不同手势状态之间的转变来控制设备。有益的是,所述方法只利用单个 HD 照相机获得作为输入的视频序列,同时能够以较低的计算复杂性获得实时性能。

[0012] 本发明的系统和方法不需要任何昂贵和特殊的传感器。通过利用肤色模型和运

动,所述系统能够在没有深度图的情况下,从背景中分割手部区域。本发明的手势组用户友好(易于记住),足以覆盖电视机控制的基本功能。

[0013] 本发明实用地实现电视机的手势控制,使实时电视机特征成为可能。另外,就本发明的系统来说,不需要使用两部照相机或者一部 3D 照相机。

[0014] 在说明书的以下各个部分中,将阐明本发明的其它各个方面,其中,详细说明用于充分公开本发明的优选实施例,而不是对本发明加以限制。

## 附图说明

[0015] 参考仅仅用于举例说明的以下附图,可更充分地理解本发明:

[0016] 图 1 是按照本发明的手势识别系统的示意图。

[0017] 图 2 是图解说明图 1 的手势识别模块的组件概况的示意图。

[0018] 图 3 是按照本发明的分割模块的示意图。

[0019] 图 4 是按照本发明,应用于图像的用于前景 / 背景分离的限界框的例示。

[0020] 图 5A-5E 图解说明按照本发明,关于不同手势的成功手部分割结果的图像。

[0021] 图 6A-E 图解说明非最佳手部分割的图像。

[0022] 图 7 图解说明按照本发明的分割图像帧的二元手部掩膜和计算的参数。

[0023] 图 8 图解说明为了得出中心 - 轮廓距离而评估的图 7 的掩膜。

[0024] 图 9 图解说明按照本发明的用于电视机控制的例证手势组。

[0025] 图 10 图解说明按照本发明,可用于特定操作的序列手势组。

[0026] 图 11 表示应用于图像的手部限界框的相邻限界框。

[0027] 图 12 表示根据分割模型的结果,转换成指示符矩阵的手部图像。

[0028] 图 13 图解说明遮挡发生的图像和相应的分割模型。

[0029] 图 14 图解说明遮挡发生的另一个图像和相应的分割模型。

[0030] 图 15 图解说明脱离遮挡的手部的图像和分割模型。

## 具体实施方式

[0031] 更具体地参考附图,出于举例说明的目的,下面用图 1- 图 15 中所示的设备具体体现本发明。要理解设备可在结构方面,以及在各个部分的细节方面发生变化,并且方法可在具体步骤和顺序方面发生变化,而不脱离这里公开的基本原理。

[0032] 图 1 图解说明本发明的手势识别系统 10 的示意图。系统 10 包括传感器 12(例如, HD 格式照相机等),其输出耦接到接收传感器输入并运行手势识别软件模块 18 的处理器 14。手势识别模块 18 分析来自传感器 12 的输入,并提供发送给设备控制应用程序或者用户界面软件 16 的命令或操作,所述设备控制应用程序或者用户界面软件 16 可用于控制音量、频道、节目或者设备的其它特征。要理解设备控制应用程序 16 可被配置成控制电视机、机顶盒、计算机或者其它组件。

[0033] 图 2 关于系统 10 的操作,图解说明手势识别模块 18 的基本组件。手势识别模块 18 一般包含 3 个阶段。在第一阶段中,在分割模块 20 和跟踪模块 22 处理来自设备 12 的输入或拍摄的图像,从而从背景中分割手部区域,以便跟踪。

[0034] 在第二阶段中,系统 10 执行特征提取模块 24,以便从检测的手部区域中提取必要

的特征。

[0035] 在第三阶段中,通过把提取的特征输入用于姿势 / 手势识别 26 的分类器,系统 10 执行手势识别模块 18。手势识别模块 18 的结果随后被输出给特定的应用程序 16,以便控制设备。

[0036] 阶段 1 :手部分割

[0037] 图像分割一般包括向图像中的像素分配标记,以致具有相同标记的像素共有某些视觉特性。本发明的系统 10 最好利用基于肤色的方法。这种方法以在每个单独的图像内,肤色始终如一(例如,数值不存在重大变化)的假定为基础。这种结构把色度分量和亮度分量分开,以降低对光变化的敏感性。这种方法为低计算复杂性和实时性能创造条件。在优选结构中,来自设备 12 的输入图像为 YCbCr444 平面颜色格式。

[0038] 如图 3 中图解所示,分割模块 20 一般包含 3 个步骤:包含前景 / 背景分离的第一步骤 30,包含模型训练的第二步骤 32,和包含像素分类的第三步骤 34。

[0039] 图 4 图解说明检测手部 44 位置和大小的前景 / 背景分离步骤 30。在检测到手部 44 之后,输出手部限界框 48。最好利用触发模块(未示出)触发本发明的系统 10 启动(例如,通过用户举起手部 44)。限界框 48 围绕前景区域 40,把背景区 42 定义为被包围在外框或者搜索区 46 中的周围图像。这里,搜索区长度被表示成 3 倍于限界框 48 的长度 L,以覆盖手部 44 的运动。不过,可以理解其它比率也是适用的。

[0040] 在模型训练步骤 32 中,最好利用 K-均值聚类(clustering)训练在手部限界框 48 内的图像 40 的 3-分量高斯混合模型(最好只使用 Cb,Cr 颜色通道)。K-均值聚类还用于训练在背景区域 42 内的图像的 3-分量高斯混合模型(只使用 Cb,Cr 颜色通道)。尽管可以使用其它方法(例如,基于直方图的方法),不过优选 K-均值算法,因为其它方法通常较慢,并且更加计算密集。

[0041] K-均值算法是用于把图像分成 K 个聚类的迭代技术。K-均值算法通常包括 1) 随机地或者根据某种试探,挑选 K 个聚类中心,2) 把图像中的每个像素分配给使该像素和聚类中心之间的距离最小化的聚类,3) 通过求聚类中的所有像素的平均值,重新计算聚类中心,和 4) 重复步骤 2 和 3,直到达到收敛为止(例如,没有像素改变聚类)。通常,所述距离是像素和聚类中心之间的平方差或者绝对值差。所述差值可以基于像素颜色、亮度、纹理和位置,或者这些因素的加权组合。

[0042] 本发明的系统 10 最好利用基于肤色的方法。这种方法以在每个单独的图像内,肤色始终如一(例如,数值不存在重大变化)的假定为基础。这种结构对光变化敏感,可通过把色度分量和亮度分量分开,加以改善。这种方法为低计算复杂性和实时性能创造条件。在优选结构中,来自设备 12 的输入图像为 YCbCr444 平面颜色格式。

[0043] 为了修正前景 40 颜色模型,除去不在一般皮肤模型的范围( $77 \leq Cb \leq 127$ ,  $133 \leq Cr \leq 173$ )内的所有分量。

[0044] 如果 3 个分量都在(或者都不在)皮肤模型范围之内,那么我们根据前景高斯模型,确定最接近任意背景分量的分量,随后除去该分量。剩余的分量随后被用于表示前景区域 40。

[0045] 最好利用 GMM 和 Bayesian 决策规则,进行像素分类步骤 34。对于每个当前的手部图像,我们利用前一个图像的不同部分作为基准。满足以下标准(关于运动的等式 1 和

2, 和关于肤色的等式 3) 任意之一的像素被视为手部像素 (否则, 像素被分类为背景像素) (以下公式中, hand 表示手, background 表示背景):

[0046] 1. 运动和肤色:

$$[0047] \quad |Cb(x, y) - Cb_{ref}(x, y)| > 5 \text{ 或 } |Cr(x, y) - Cr_{ref}(x, y)| > 5 \quad \text{等式 1}$$

[0048] 和

$$[0049] \quad 2 * P_{hand} * P(c | hand) > (1 - P_{hand}) * P(c | background) \quad \text{等式 2}$$

[0050] 2. 肤色:

$$[0051] \quad P_{hand} * P(c | hand) > (1 - P_{hand}) * P(c | background) \quad \text{等式 3}$$

[0052] 像素分类步骤 34 随后最好继之以形态学膨胀和侵蚀算子, 以修正获得的二元手部掩膜。

[0053] 如下利用等式 4、5 和 6, 计算像素分类步骤 34 中的概率:

$$[0054] \quad P(c | hand) = \sum_{k=1}^K p_k g_{hand}(c; \mu_k, \sigma_k) \quad \text{等式 4}$$

[0055] 其中  $P_{hand}$  是先前的手部区域与背景区域 42 之比,  $P(c | hand)$  是像素  $c$  属于前景 (手部) 区域 40 的概率,  $g_{hand}$  是像素  $c$  属于前景高斯混合模型的第  $k$  个分量的概率,  $p_k$  是第  $k$  个分量的先验概率 (在模型训练步骤 32 中训练  $p_k$ , 均值  $\mu_k$  和协方差矩阵  $\sigma_k$ )。

$$[0056] \quad P(c | background) = \sum_{k=1}^K p_k g_{background}(c; \mu_k, \sigma_k) \quad \text{等式 5}$$

[0057] 其中  $P(c | background)$  是像素  $c$  属于背景区域的概率,  $g_{background}$  是像素  $c$  属于背景高斯混合模型的第  $k$  个分量的概率,  $p_k$  是第  $k$  个分量的先验概率 (在模型训练步骤 32 中训练  $p_k$ , 均值  $\mu_k$  和协方差矩阵  $\sigma_k$ )。

[0058] 利用等式 6, 计算像素  $c$  属于 GMM 分量的概率:

$$[0059] \quad g(c; \mu_k, \sigma_k) = \frac{1}{(\sqrt{2\pi}\sigma_k)^2} e^{-\frac{1}{2} \left( \frac{\|c - \mu_k\|}{\sigma_k} \right)^2} \quad \text{等式 6}$$

[0060] 图 5A-5E 图解说明按照本发明, 关于不同手势的成功手部分割结果。对应地, 图 6A-E 中的图像图解说明非最佳的手部分割, 其中某些手势的图像具有局部特征, 缺少的, 不正确的和 / 或不连续的解剖体, 如包围的区域中所示。

[0061] 跟踪步骤 22 跟踪作为观察点的分割图像的手掌中心。利用下面详细说明了基于图像的方法, 获得手掌中心。

[0062] 跟踪步骤把 Kalman 滤波器用于下一帧的实际手部位置和速度的预测和计算 (根据对当前帧的观察)。

[0063] 等式 7 和 8 中表示了利用较少计算的例证 Kalman 滤波器:

$$[0064] \quad X_k = F_k X_{k-1} + B_k U_k + W_k \quad \text{等式 7}$$

[0065] 其中  $F_k$  是应用于前一个状态  $X_{k-1}$  的状态转变模型,  $B_k$  是应用于控制向量  $U_k$  的控制输入模型,  $W_k$  是假定从具有协方差  $Q_k$  的零均值多元正态分布得到的处理噪声。

$$[0066] \quad Z_k = H_k X_k + V_k, \quad V_k \sim N(0, R_k) \quad \text{等式 8}$$



[0067] 其中  $H_k$  是把真实状态空间映射到观察空间的观察模型,  $V_k$  是被假定为具有协方差  $R_k$  的零均值高斯白噪声的观察噪声。

[0068] 随后输出每个图像中的手部位置(例如,手掌中心的2D位置),并保存为输出结构的一部分。

[0069] 阶段2:特征提取

[0070] 图7-8图解说明特征提取模块24的第二阶段的操作。

[0071] 图7图解说明每个分割的图像帧的二元手部掩膜60。手部掩膜60包括根据原始掩膜和侵蚀掩膜(未示出)之间的差异产生的提取轮廓62,和骨架64。侵蚀掩膜是利用现有技术中一般可用的侵蚀算法获得的,所述侵蚀算法使原始掩膜细化(thin)一个像素。根据把从轮廓62获得的二元形状转换成1像素宽的线条的形态学细化算法,生成骨架64。所述方法可包含本领域中通常已知的任何细化算法,不过通常涉及反复删除轮廓形状62内的像素,以在不缩短轮廓形状62或者分裂轮廓形状62的情况下,收缩轮廓形状62。

[0072] 随后相对于骨架64分析提取的轮廓62,从而获得手掌中心  $C_p$ , 手掌中心  $C_p$  被定义和计算为骨架64上,距手部轮廓的距离为最大距离  $r$  的点。

[0073] 随后手掌底部  $B_p$  被计算为骨架64上,距离手掌中心  $C_p$ (在手掌中心  $C_p$  下面)  $r$  个像素的点。

[0074] 在对于每个手部掩膜60确定了手掌中心  $C_p$  和手掌底部  $B_p$  点之后,方向向量  $V_b$  被计算为通过手掌中心  $C_p$  和手掌底部  $B_p$  这两个点的线条。

[0075] 随后计算在方向向量  $V_b$  的  $\pm 60^\circ$  范围内,具有距手掌中心  $C_p$ (或者作为替换方案,手掌底部  $B_p$ ) 的最大距离  $D$  的轮廓点66(通常是中指指尖)。根据平均解剖学,通常使用  $\pm 60^\circ$  的范围。不过,可以相应地改变该范围。

[0076] 阶段3:手势识别

[0077] 对于手势识别模块26,通过比较距离  $D$  和阈值  $T$ ,随后把每个手部图像分类成“张开”或“握紧”。阈值  $T$  是借助校准步骤获得的,在校准步骤中,分割前5个“张开的手部”图像,获得距离  $D$  和手掌半径  $r$  的值。随后利用平均值  $AVE(D+1.5r)/2$ ,计算阈值。

[0078] 在每个视频中,具有5帧大小的滑动窗口将识别“张开的”或者“握紧的”手势。

[0079] “张开”:如果在最近5帧内,存在3个或者更多的张开姿势,那么认为当前姿势是“张开”。

[0080] “握紧”:如果在最近5帧内,存在3个或者更多的握紧姿势,那么认为当前姿势是“握紧”。

[0081] 图9图解说明本发明的手势识别模块26的某些关键功能。“张开”姿势和“握紧”姿势之间的转变可用于指示何时触发了手势“选择”或者“抓取”。

[0082] 例如,为了触发手势识别系统初始化,可在传感器12的视场内的区域中举起手部44,并以张开手部姿势保持1-2秒,如步骤70中所示。

[0083] 一旦系统被初始化,在步骤72,就可通过跟踪自由手部运动(例如,手掌中心),把张开的手部用于控制光标(在操作的电视机或设备(例如,机顶盒等)的监视器屏幕上,光标是可见的)。

[0084] 通过利用手势与不同定时的组合,可以执行特殊的命令,例如,抓取/附着操作74或者选择/检查操作76。图10图解说明可用于抓取/附着操作74或者选择/检查操作76

的序列。第一步骤 90 包括张开手部手势 44, 该手势可把光标指引到期望的位置 (例如, 用户界面屏幕上的位置), 以便进行检查或选择。借助握紧拳头姿势 80, 手指向前手势 82 或者手指捏紧手势 84, 可以启动第二步骤 92, 以选择或检查期望的特征。一旦已选择, 第三步骤 94 中的张开手掌姿势 44 恢复光标的自由移动。

[0085] 类似地进行抓取 / 附着操作 74, 不过, 用户应该保持步骤 2 中的手势 80、82 或 84 之一, 等待物品附着上。

[0086] 现在参考图 11-15, 本发明的系统 10 最好包括手部 - 面部遮挡程序, 以处理当手部 44 越过或者非常接近面部 102 时的情况。

[0087] 遮挡情况通常受以下假定限制: 1) 手部具有远远大于面部的运动, 2) 不存在靠近面部的任何其它的肤色团块, 在遮挡期间, 不存在手势的变化, 3) 在遮挡期间, 手部在  $y$  方向没有任何较大的运动。

[0088] 如图 11 中所示, 对于每个当前的手部限界框 48, 通过利用两个矩形 94 和 96, 生成在框 48 的左右两侧的相邻区域。每个矩形 94 和 96 高度为  $1.5L$ , 宽度为  $L$ , 其中  $L$  是限界框 48 长度。框 48、94 和 96 的底部被对齐, 以致左右相邻框 94 和 96 延伸高过框 48  $1/2L$ 。

[0089] 相邻区域 94, 96 中的所有像素被分类为皮肤像素或非皮肤像素 (通过利用先前训练的高斯混合模型)。随后计数每个正方形区域中的皮肤像素的数目。如果所述数目大于阈值 (例如, 在校准步骤中计算的张开的手部的  $2/3$ ), 那么该正方形被认为包含会导致遮挡的对象。

[0090] 如图 12 中所示, 根据分割模型 104 的结果, 每个手部图像 100 被转换成指示符矩阵 106。包含对象的框被赋值“1”, 否则被赋值“0”。如图 12 中所示, 左侧的相邻框 94 不满足阈值皮肤像素, 从而被赋值“0”, 而手部限界框 48 和右侧相邻框 96 满足阈值像素计数, 从而被赋值“1”。

[0091] 之后, 计算并记录对象  $C_0$  的质心。检测每对相邻的指示符矩阵 (沿着时间轴)。具有  $1 \rightarrow 0$  的转变的任何位置指示以下之一: 1) 附近的对象 (面部 102) 合并于手部 44, 或者 2) 手部 44 正在离开对象 (例如, 面部 102)。

[0092] 测量手部的当前位置与先前记录的对象质心  $C_0$  的距离, 以确定手部 44 是正在接触该对象, 还是正在远离该对象 (例如, 如果所述距离小于指定百分率 (例如, 60-75%) 的限界框长度  $L$ , 那么发生遮挡)。

[0093] 图 13 图解说明遮挡发生的图像 112 和相应的分割模型 114。如果发生遮挡, 那么定义接近于当前质心  $C_c$  的位置的备用区域 110 (具有与质心相同的  $y$ )。备用区域 110 的大小与手部限界框 48 相同, 例如长度  $L$ 。

[0094] 图 14 图解说明遮挡发生的另一个图像 116, 和相应的分割模型 118。在该遮挡期间, 整个二元掩膜 120 (手部 44 加面部 102) 的质心  $C_c$  被用作预测位置。

[0095] 图 15 图解说明脱离遮挡 (例如, 在备用区域 110 中检测到对象) 的手部 44 的图像 122 和分割模型 124。对象的质心  $C_c$  随后被用作当前手部位置, 并重新开始运动跟踪。

[0096] 总之, 本发明的系统 10 和方法应用与运动信息结合的基于肤色的方法来实现实时分割。利用 Kalman 滤波器跟踪手部的质心。计算提取的手部掩膜的手掌中心, 手掌底部, 以及从手掌中心到轮廓的最大距离。随后比较计算的距离和阈值, 以确定当前姿势是“张开”还是“握紧”。在优选实施例中, “张开”姿势和“握紧”姿势之间的转变确定当前手势是

处于“选择”状态还是“抓取”状态。

[0097] 参考按照本发明的实施例的方法和系统的流程图,说明了本发明的实施例。这些方法和系统也可被实现成计算机程序产品。在这点上,流程图的每个方框或步骤,和流程图中的各个方框(和/或各个步骤)的组合可用各种装置,比如硬件、固件和/或包括用计算机可读程序代码逻辑具体体现的一个或多个计算机程序指令的软件实现。可以理解,任何这样的计算机程序指令可被加载到计算机,包括(但不限于)通用计算机或专用计算机,或者其它可编程处理设备,从而产生机器,以致在计算机或其它可编程处理设备上执行的计算机程序指令产生实现现在一个或多个流程图的一个或多个方框中指定的功能的装置。

[0098] 因而,流程图的方框支持完成指定功能的装置的组合,实现指定功能的步骤的组合,和实现指定功能的计算机程序指令,比如用计算机可读程序代码逻辑装置具体体现的计算机程序指令。另外显然流程图的每个方框,和流程图中的各个方框的组合可用完成指定功能或步骤的基于专用硬件的计算机系统,或者专用硬件和计算机可读程序代码逻辑装置的组合实现。

[0099] 此外,这些计算机程序指令,比如用计算机可读程序代码逻辑具体体现的计算机程序指令也可被保存在计算机可读存储器中,所述计算机程序指令可指令计算机或其它可编程处理设备按照特定方式工作,以致保存在计算机可读存储器中的指令产生制品,所述制品包括实现现在一个或多个流程图的一个或多个方框中指定的功能的指令装置。计算机程序指令也可被加载到计算机或者其它可编程处理设备上,使得在计算机或其它可编程处理设备上执行一系列的操作步骤,从而产生计算机实现的处理,以致在计算机或其它可编程处理设备上执行的指令提供实现现在一个或多个流程图的一个或多个方框中指定的功能的步骤。

[0100] 于是可以看出,本发明包括以下发明实施例:

[0101] 1. 一种用于遥控设备的手势识别系统,包括:在所述设备附近的位置,捕捉用户的手部的视频数据的传感器;处理捕捉的所述手部的视频数据的处理器;可在所述处理器上运行,以执行以下步骤的程序:根据肤色,分割捕捉的视频中的每个手部图像;提取分割的手部图像的一个或多个参数;跟踪手部图像中的一个或多个参数中的至少一个参数;根据所述一个或多个参数,把手势分类成“张开”或“握紧”;和根据所述手势的识别和所述一个或多个参数的跟踪,操纵所述设备。

[0102] 2、如在实施例 1 中所述的系统,其中分割每个手部图像包括:把与所述手部相关的前景区域和围绕所述前景区域的背景区域分开;训练背景区域和前景区域的 3 分量混合模型;和根据运动和肤色中的一个或多个,对图像像素分类。

[0103] 3、如在实施例 1 中所述的系统,其中只利用 Cb 和 Cr 颜色通道,分割图像。

[0104] 4、如在实施例 1 中所述的系统,其中提取一个或多个参数包括:根据提取的手部的轮廓和骨架,获得用户的手部的手掌中心位置。

[0105] 5、如在实施例 4 中所述的系统,其中跟踪一个或多个参数包括利用 Kalman 滤波器,跟踪手掌中心位置。

[0106] 6、如在实施例 4 中所述的系统,其中提取一个或多个参数还包括:获得手掌底部位置;和测量手掌中心位置与提取的图像轮廓上的最远轮廓点之间的距离;所述最远轮廓点在通过手掌底部位置和手掌中心位置的向量的预定角度范围内。

[0107] 7、如在实施例 6 中所述的系统,其中把手势分类成“张开”或“握紧”包括:比较到最远轮廓点的距离和阈值;如果到最远轮廓点的距离超出阈值,那么把图像分类成“张开”;如果到最远轮廓点的距离低于阈值,那么把图像分类成“握紧”。

[0108] 8、如在实施例 1 中所述的系统,其中设备包含与显示器耦接的用户界面,所述用户界面是在所述显示器上输出的;其中操纵所述设备包括根据所述手势的识别,和所述一个或多个参数的跟踪,向用户界面发送命令。

[0109] 9、如在实施例 8 中所述的系统,其中“张开”手势和“握紧”手势之间的转变被用于确定“选择”或“抓取”命令是否被发送给用户界面。

[0110] 10、如在实施例 9 中所述的系统,其中张开手势的跟踪被用于确定光标在用户界面上的位置。

[0111] 11、一种用于遥控设备的手势识别系统,所述设备具有在显示器上显现的用户界面,所述系统包括:在所述设备附近的位置,捕捉用户的手部的视频数据的传感器;处理捕捉的所述手部的视频数据的处理器;可在所述处理器上运行,以执行以下步骤的程序:根据肤色,分割捕捉的视频中的每个手部图像;提取分割的手部图像的一个或多个参数;跟踪手部图像中的所述一个或多个参数;根据所述一个或多个参数,把手势分类成“张开”或“握紧”;和根据所述手势的识别和所述一个或多个参数的跟踪,操纵所述设备;其中操纵所述设备包括根据所述手势的识别,和跟踪所述一个或多个参数的位置,向用户界面发送命令。

[0112] 12、如在实施例 11 中所述的系统,其中分割每个手部图像包括:把与所述手部相关的前景区域和围绕所述前景区域的背景区域分开;训练背景区域和前景区域的只包含 Cb 和 Cr 颜色通道的 3 分量混合模型;和根据运动和肤色中的一个或多个,对图像像素分类。

[0113] 13、如在实施例 11 中所述的系统,其中提取一个或多个参数包括:根据提取的手部的轮廓和骨架,获得用户的手部的手掌中心位置;根据手掌中心位置,获得手部的手掌底部位置;和测量手掌中心位置与提取的图像轮廓上的最远轮廓点之间的距离;所述最远轮廓点在通过手掌底部位置和手掌中心位置的向量的一定角度范围内。

[0114] 14、如在实施例 13 中所述的系统,其中跟踪所述一个或多个参数包括利用 Kalman 滤波器,跟踪手掌中心位置。

[0115] 15、如在实施例 13 中所述的系统,其中把手势分类成“张开”或“握紧”包括:比较到最远轮廓点的距离和阈值;如果到最远轮廓点的距离超出阈值,那么把图像分类成“张开”;如果到最远轮廓点的距离低于阈值,那么把图像分类成“握紧”。

[0116] 16、如在实施例 15 中所述的系统,其中“张开”手势和“握紧”手势之间的转变被用于确定“选择”或“抓取”命令是否被发送给用户界面。

[0117] 17、如在实施例 16 中所述的系统,其中张开手势的跟踪被用于确定光标在用户界面上的位置。

[0118] 18、一种利用手势,遥控设备的方法,所述设备具有在显示器上显现的用户界面,所述方法包括:用在所述设备附近位置的传感器,捕捉用户的手部的视频;根据肤色,分割捕捉的视频中的每个手部图像;提取分割的手部图像的一个或多个参数;跟踪手部图像中的所述一个或多个参数中的至少一个参数;根据所述一个或多个参数,把手势分类成“张

开”或“握紧”；根据所述手势的识别和所述一个或多个参数的跟踪，操纵所述设备；根据所述手势的识别，和所述一个或多个参数的跟踪，向用户界面发送命令；和在显示器上输出发送的命令，以便操纵所述设备。

[0119] 19、如在实施例 18 中所述的方法，其中提取一个或多个参数包括：根据提取的手部的轮廓和骨架，获得用户的手部的手掌中心位置；根据手掌中心位置，获得手部的手掌底部位置；和测量手掌中心位置与提取的图像轮廓上的最远轮廓点之间的距离；所述最远轮廓点在通过手掌底部位置和手掌中心位置的向量的一定角度范围内，其中把手势分类成“张开”或“握紧”包括：比较到最远轮廓点的距离和阈值；如果到最远轮廓点的距离超出阈值，那么把图像分类成“张开”；如果到最远轮廓点的距离低于阈值，那么把图像分类成“握紧”。

[0120] 20、如在实施例 19 中所述的方法，其中“张开”手势和“握紧”手势之间的转变被用于确定“选择”或“抓取”命令是否被发送给用户界面；其中张开手势的手掌中心的跟踪被用于确定光标在用户界面上的位置。

[0121] 虽然上面的说明包含许多细节，不过这些细节不应被理解为限制本发明的范围，而应被理解为仅仅提供本发明的一些目前优选的实施例的例示。于是，要理解本发明的范围完全包含对本领域的技术人员来说显而易见的其它实施例，本发明的范围因而仅仅由附加的权利要求限定，其中除非另有说明，否则对元件的单数引用并不意味“有且仅有一个”，而是“一个或多个”。本领域的普通技术人员已知的上述优选实施例的元件的所有结构、化学或功能等同物引用地明确包含于其中，被权利要求所覆盖。此外，设备或方法不必解决试图由本发明解决的每一个问题，因为它将被权利要求所覆盖。此外，本公开中的元件、组件或方法步骤都不意图贡献给公众，与是否在权利要求中明确地记载了所述元件、组件或方法步骤无关。不应按照 35U. S. C. 112, 第 6 段的规定解释权利要求的任何要素，除非该要素是利用短语“用于……的装置”明确记载的。

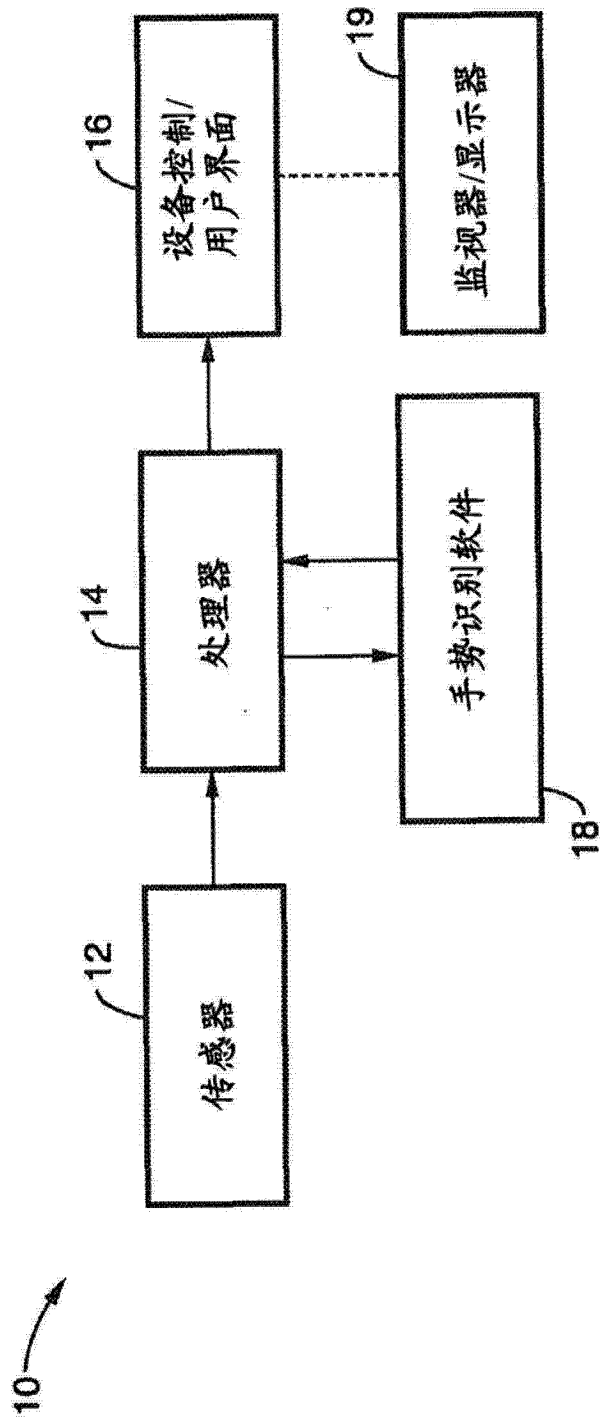


图 1

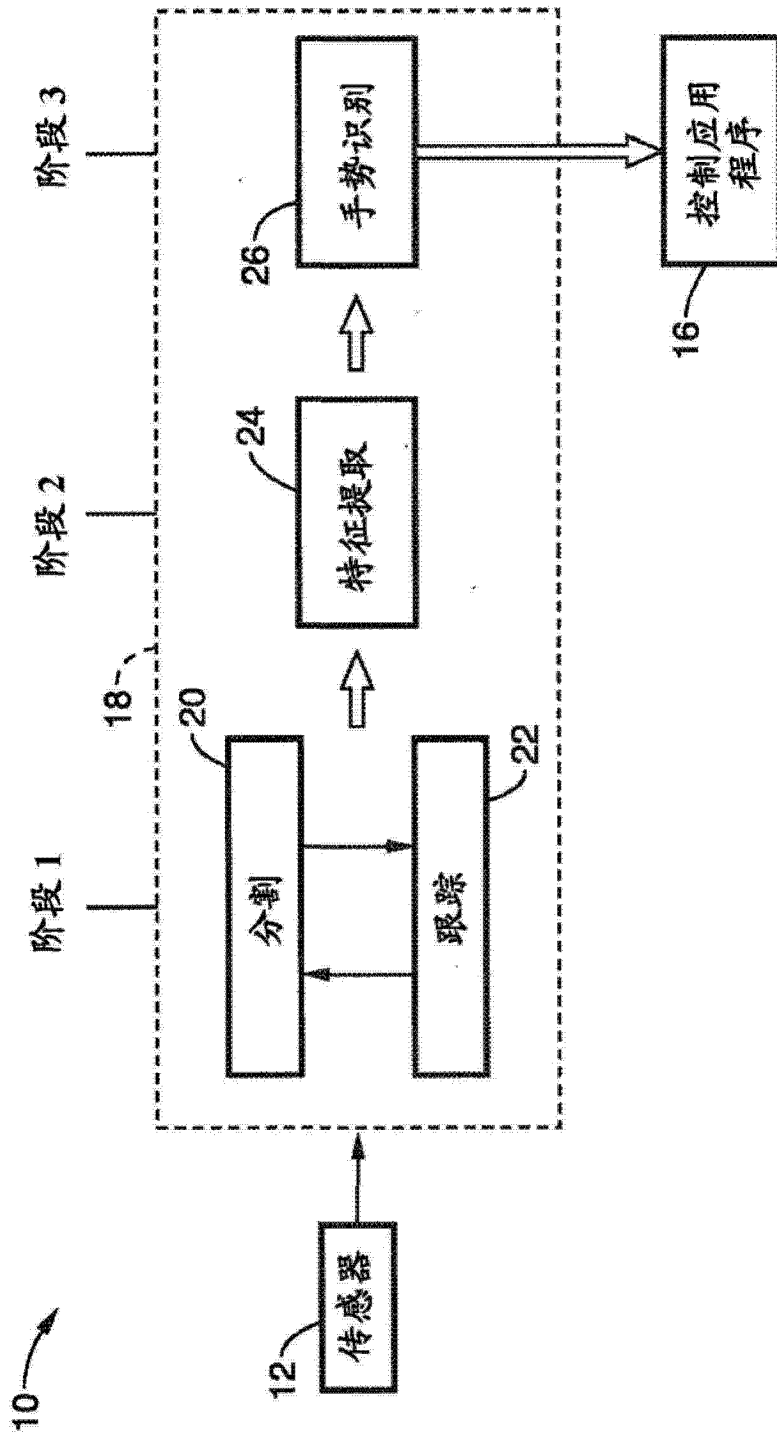


图 2

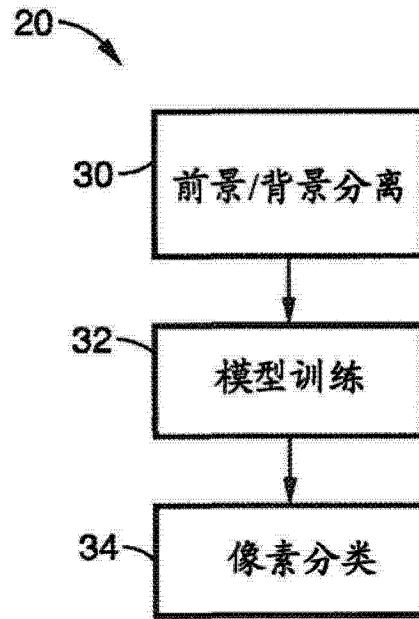


图 3

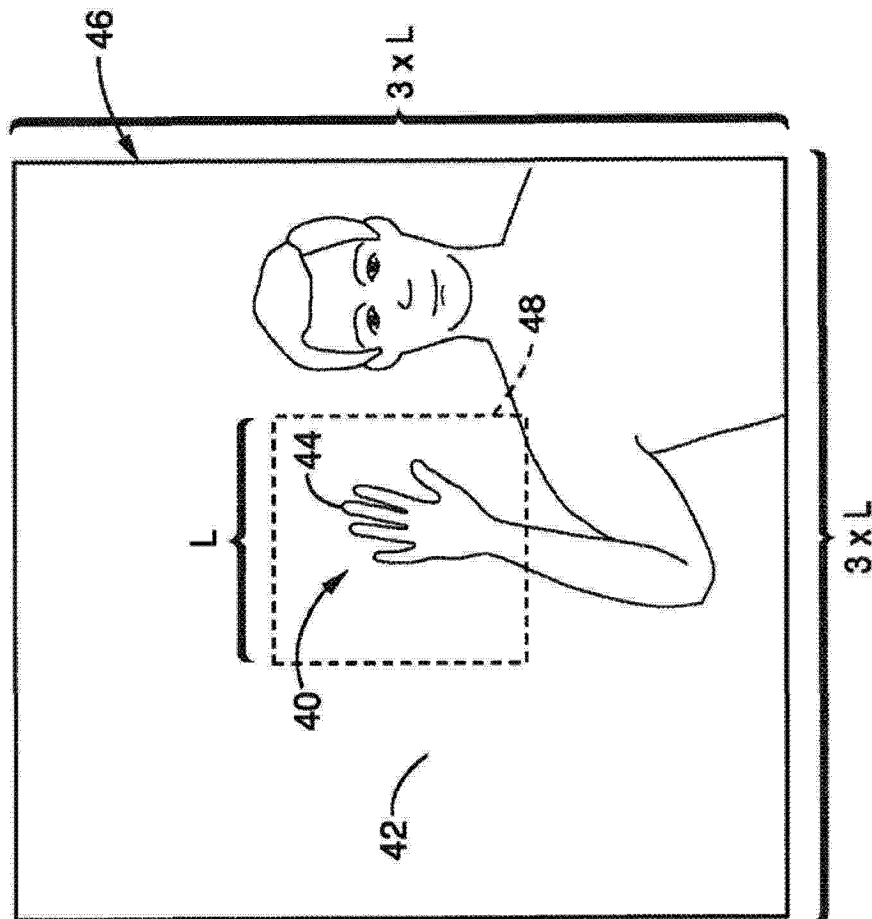


图 4



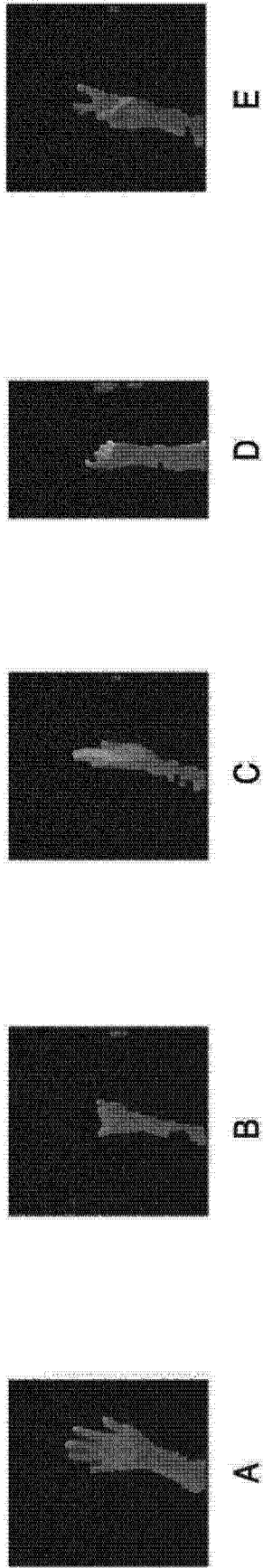


图 5

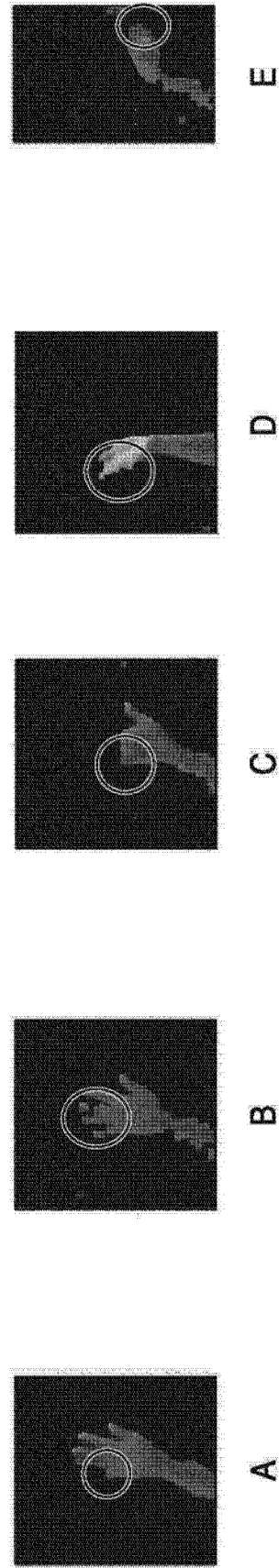


图 6

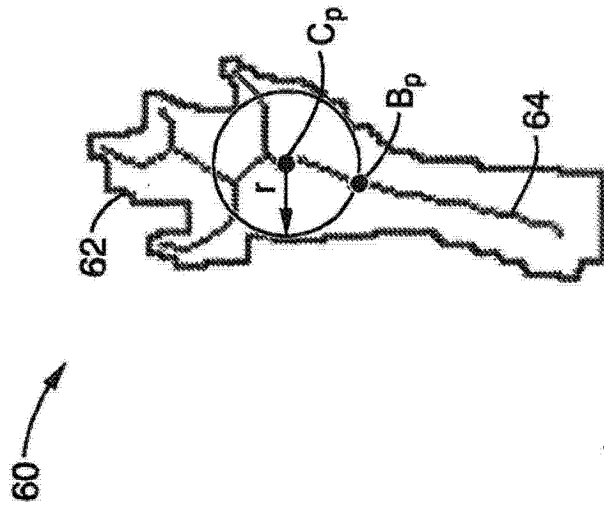


图 7

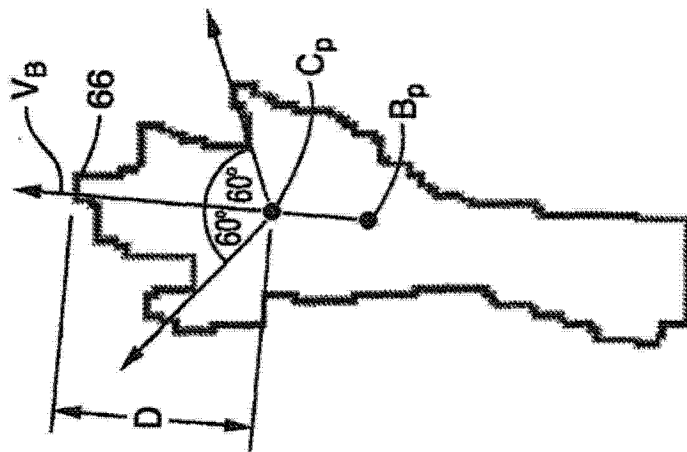


图 8

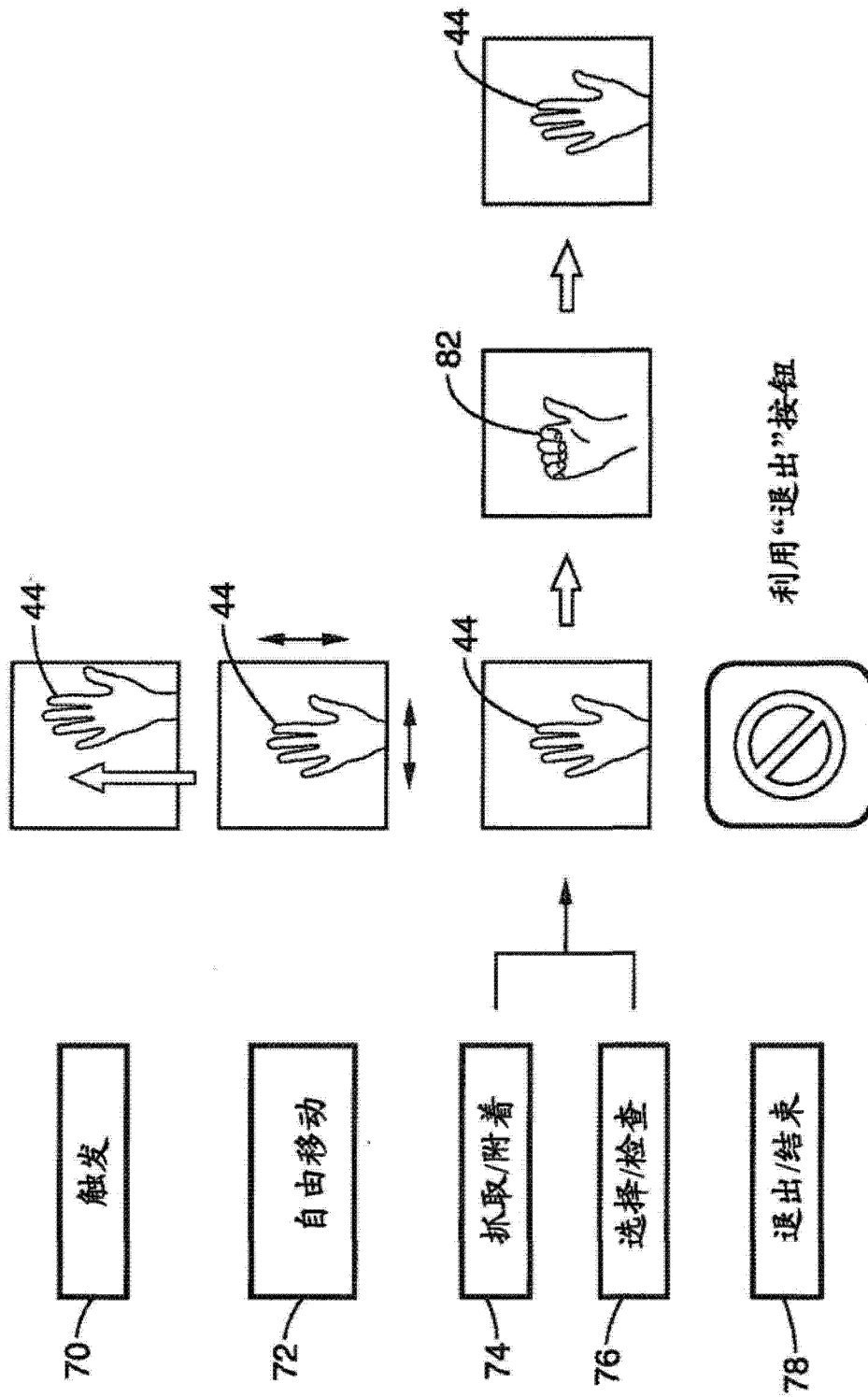


图 9

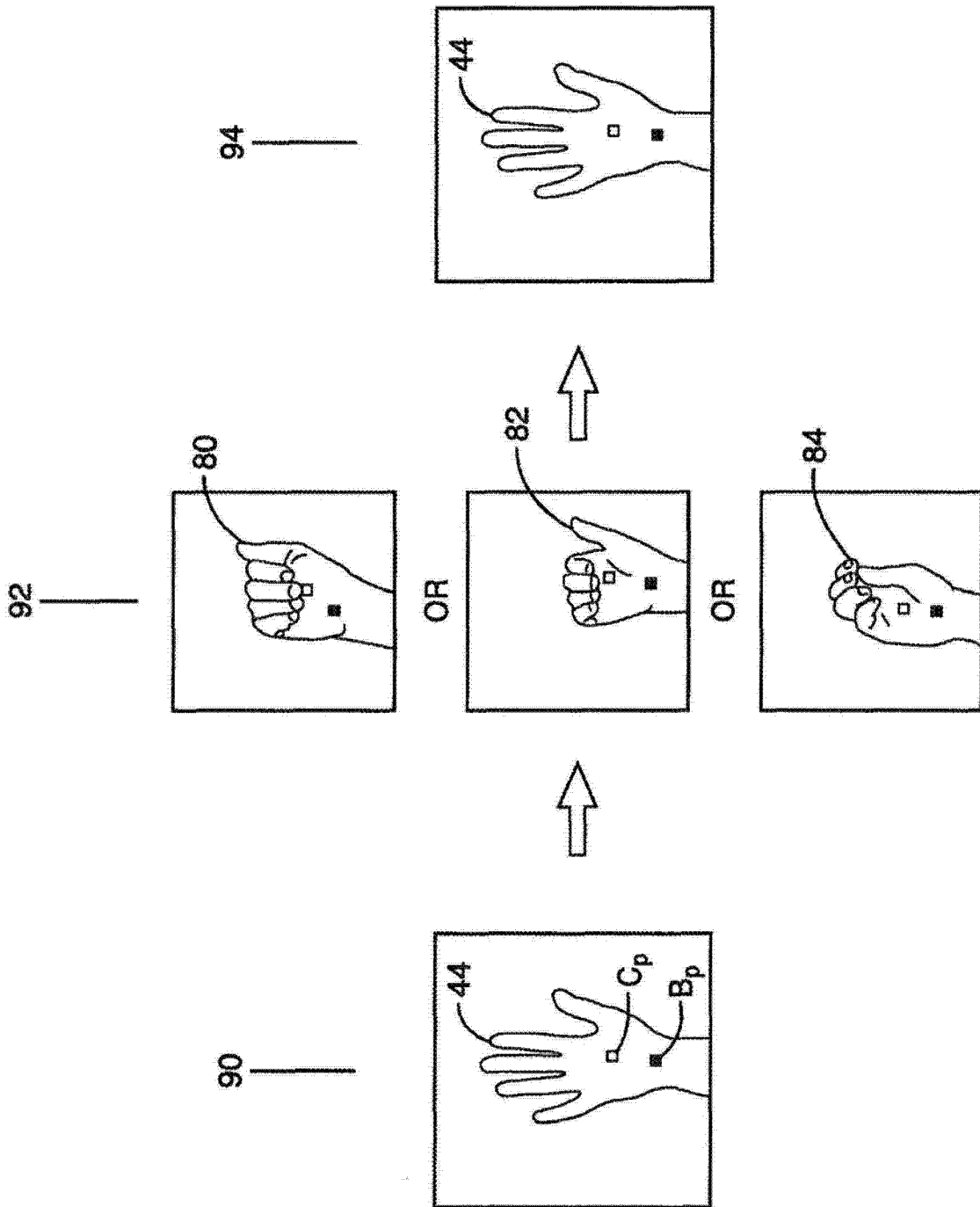


图 10

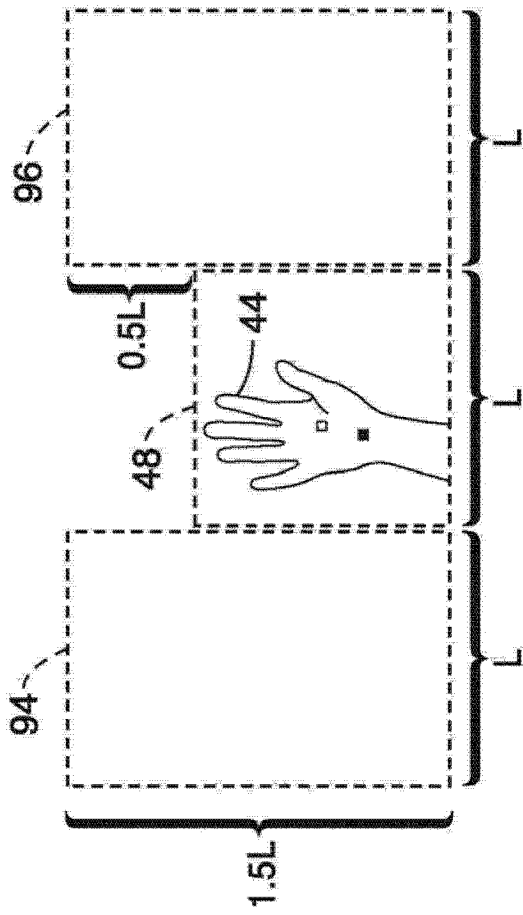


图 11

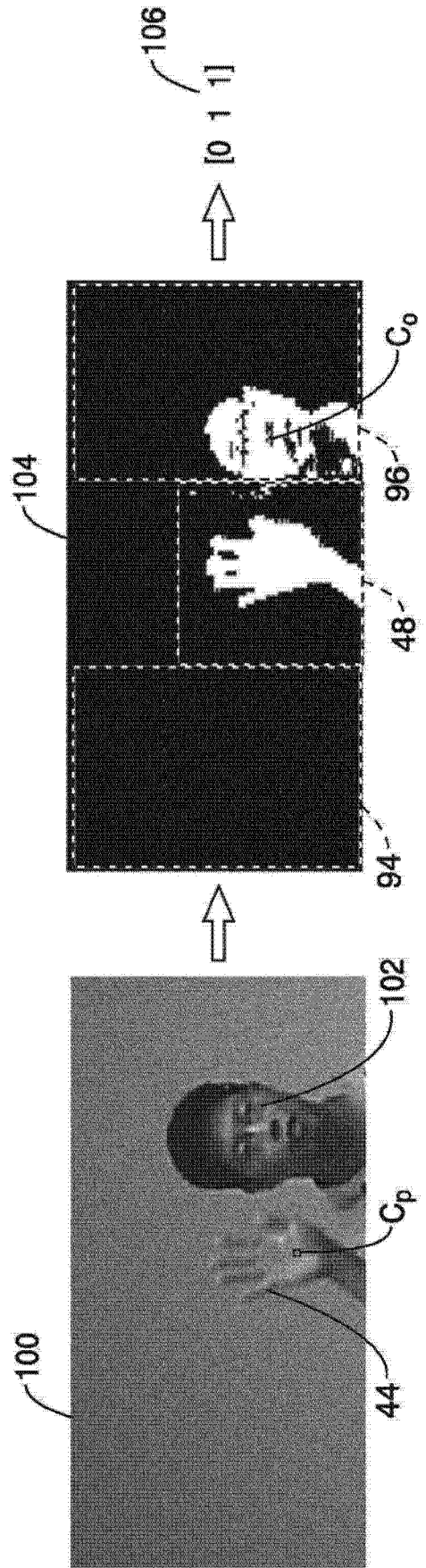


图 12

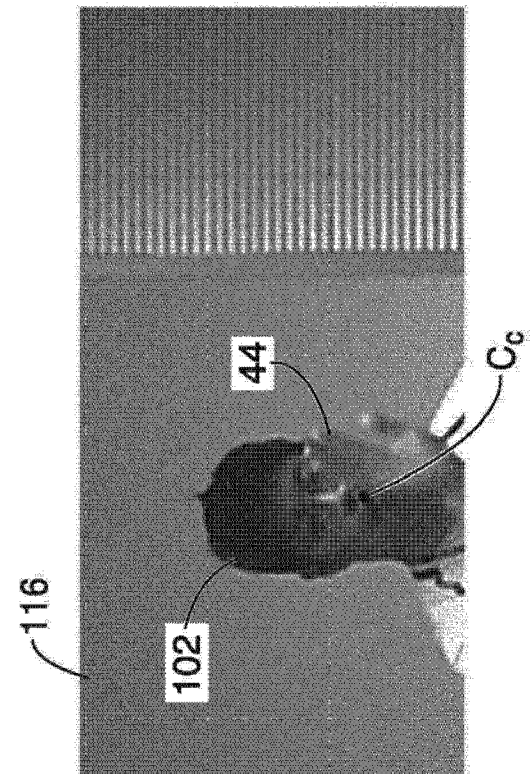
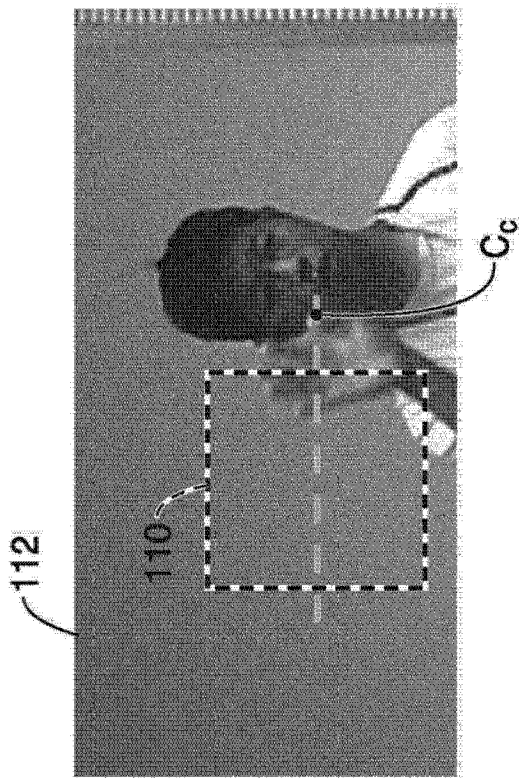
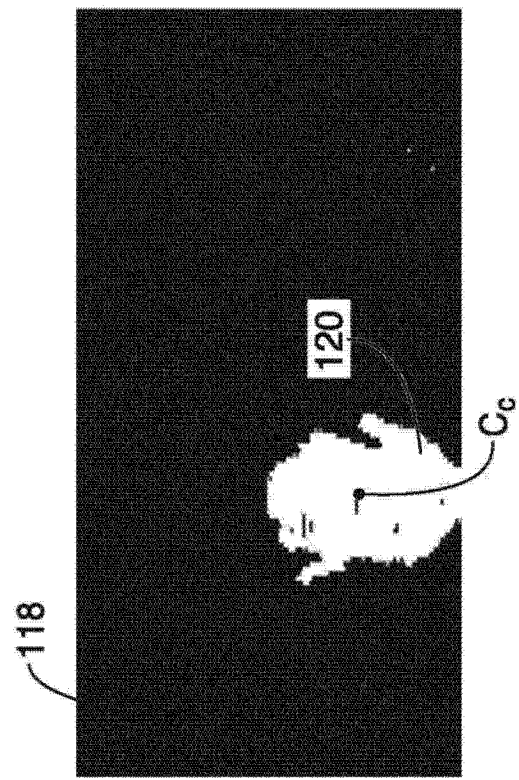
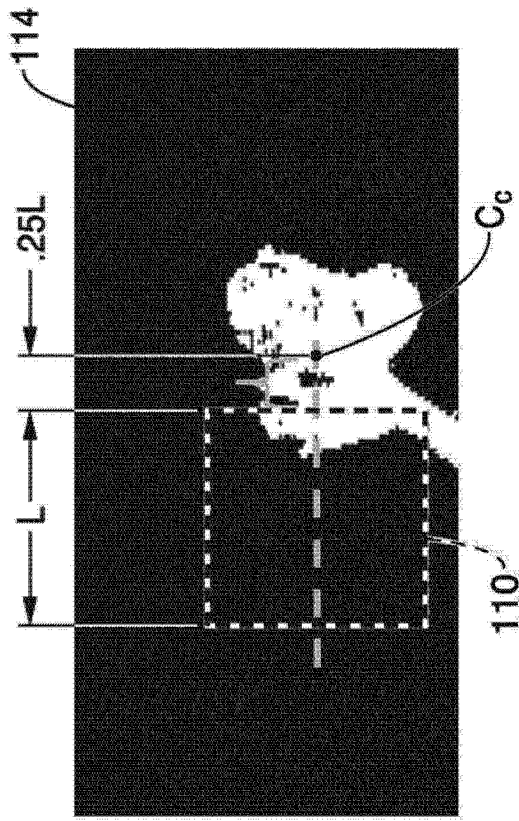


图 13

图 14

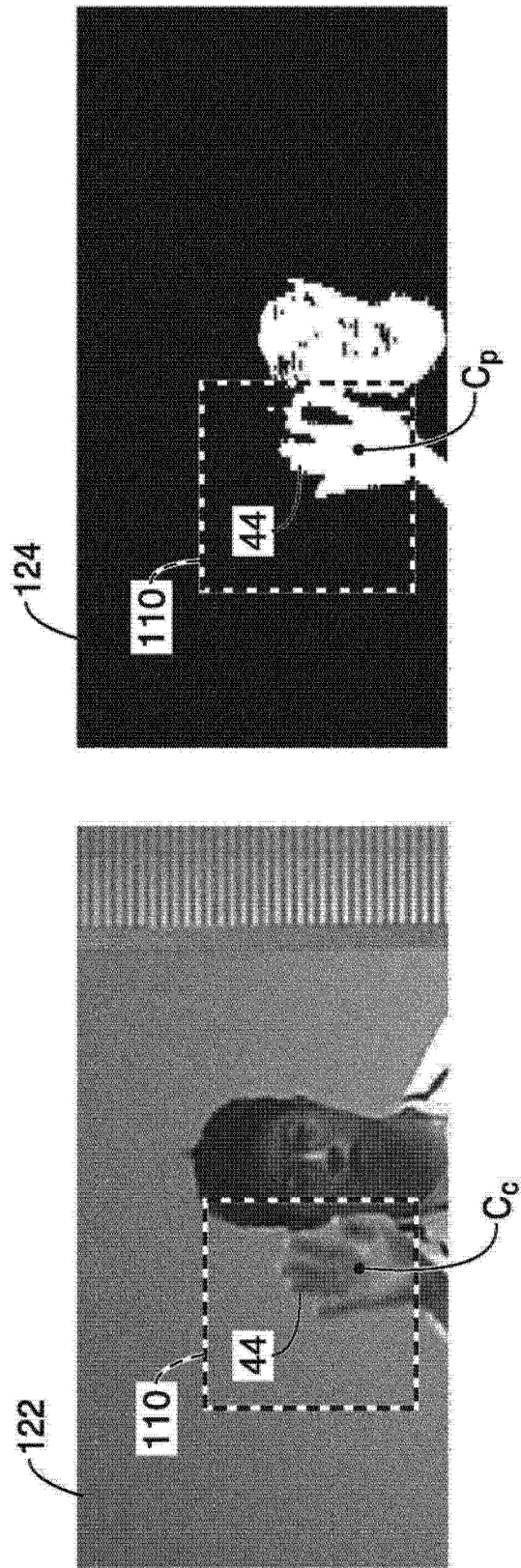


图 15