[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06K 9/00 (2006.01)
G06T 7/00 (2006.01)

[12] 发明专利说明书

专利号 ZL 03812563.3

[45] 授权公告日 2008 年 4 月 30 日

[11] 授权公告号 CN 100385448C

[22] 申请日 2003.4.4 [21] 申请号 03812563.3

[30] 优先权

[32] 2002. 4. 5 [33] US [31] 10/116,839

[86] 国际申请 PCT/IB2003/001298 2003.4.4

[87] 国际公布 WO2003/085589 英 2003.10.16

[85] 进入国家阶段日期 2004.11.30

[73] 专利权人 伊登蒂克斯公司

地址 美国明尼苏达州

[72] 发明人 B·C·德利恩

[56] 参考文献

US5561718A 1996.10.1

US5067162A 1991.11.19

US6147678A 2000.11.14

WO01/75774A1 2001.10.11

WO99/04360A1 1999.1.28

EP0962894A2 1999.12.8

US5612686A 1997.3.18

EP0932115A2 1999.7.28

EP0918300A2 1999.5.26

审查员 毛 燕

[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

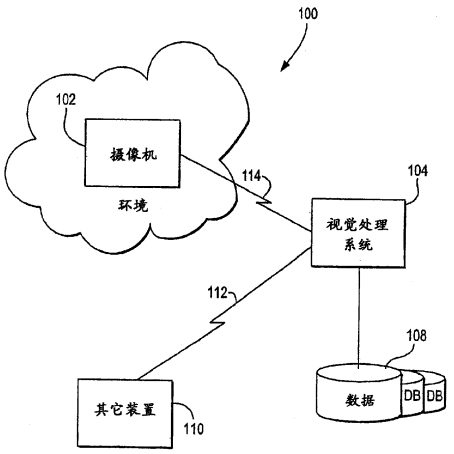
权利要求书 6 页 说明书 24 页 附图 22 页

[54] 发明名称

基于视觉的操作方法和系统

[57] 摘要

提供了一种在种类多样的环境中的基于视觉的操作系统，其中，基于计算机的系统接收视觉输入，并采取合适的动作。 一种环境是安全系统，其中，基于视觉的操作系统捕获面部图像，并以很小的假肯定概率确定出该面部是否是授权进入此环境的人的面部。



1. 一种用于确定捕获的图像是否匹配参考图像的方法，所述方法包括以下步骤:

捕获数字图像，用于与参考图像进行比较;

将所述捕获的图像分成像素组;

识别所述捕获的图像中的一组像素;

获得参考图像;

为像素组检索参考图像，该像素组与识别步骤中在捕获的图像中识别的像素组相类似;

在所述参考图像中选择与所述捕获的图像中的像素组最佳匹配的像素组;

针对至少一个另外的组像素重复上述识别、检索和选择步骤;

将所述参考图像中的选定像素组和所述捕获的图像中的像素组的相对位置进行比较；以及

确定所述相对位置随机出现的概率。

2. 如权利要求 1 所述的方法，其特征在于，所述方法还包括:

将随机匹配的概率与阈值概率比较，用于推断图像是否匹配。

3. 如权利要求 2 所述的方法，其特征在于，所述阈值概率设置成将错误推断的概率减小到任意低的数。

4. 如权利要求 1 所述的方法，其特征在于，寻找最近的像素组采用了所述参考图像中的所述像素组与所述捕获的图像中的所述选定像素组之间差异的最小平方计算。

5. 如权利要求 1 所述的方法，其特征在于还包括:

一旦捕获所述参考图像和所述捕获的图像，在完成其它步骤之前，过滤所述图像。

6. 如权利要求 5 所述的方法，其特征在于，所述过滤包括:

基于相邻像素的平均值来确定像素亮度;

将所述像素的亮度值与阈值比较;

如果所述亮度小于阈值，则将所述像素的过滤值设定为零；以及

如果所述亮度等于或大于阈值，则将所述像素的过滤值设定为1。

7. 如权利要求6所述的方法，其特征在于还包括:

通过将已知的特征定位到预定位置上，来归一化所述捕获的图像和所述参考图像。

8. 如权利要求1所述的方法，其特征在于，捕获所述图像是通过具有基于视觉的操作系统的视觉处理系统来实现的。

9. 如权利要求1所述的方法，其特征在于还包括：将所述捕获的图像与一个以上参考图像进行比较。

10. 如权利要求1所述的方法，其特征在于，比较所述捕获的图像的像素组与所述参考图像的像素组的相对位置包括:

定义所述捕获的图像的第一像素组与所述捕获的图像的第二像素组之间的第一向量;

定义所述参考图像的第一像素组与所述参考图像的第二像素组之间的第二向量；以及

计算所述第一向量与所述第二向量之间差异的绝对值。

11. 如权利要求10所述的方法，其特征在于还包括:

对于其它像素组重复所述计算。

12. 如权利要求1所述的方法，其特征在于还包括:

将随机匹配的概率与阈值概率进行比较，以推断所述图像匹配;

对于不同的参考图像重复所述步骤；以及

在得到与所述参考图像之一的匹配之前，允许不限数量的不匹配。

13. 如权利要求1所述的方法，其特征在于还包括：在确定匹配时，将所述捕获的图像添加到参考图像的数据库中，以便于以后

的匹配。

14. 如权利要求 1 所述的方法，其特征在于，所述捕获的数字图像对应于皮肤块。

15. 一种用于确定捕获的图像是否匹配参考图像的系统，它包括:

摄像机，用于捕获数字图像以与参考图像比较;

识别模块，用于将所述捕获的图像分成像素组以及用于识别所述捕获的图像中的一组像素;

选择模块，用于获得参考图像，为与所述识别模块识别的像素组相类似的像素组检索参考图像，以及在所述参考图像中选择与所述捕获的图像中的像素组最佳匹配的像素组;

比较模块，用于对至少一个更多的像素组重复上述由识别模块实施的识别步骤以及由选择模块实施的检索和选择步骤，并把所述参考图像中的选择的像素组的相对位置与所述捕获的图像中的像素组的相对位置进行比较; 和

计算器，用于确定所述相对位置随机出现的概率。

16. 如权利要求 15 所述的系统，其特征在于还包括:

匹配模块，用于将随机匹配的概率与阈值概率进行比较，用于推断所述图像匹配。

17. 如权利要求 16 所述的系统，其特征在于，所述阈值概率设置成可将错误推断的概率降到任意低的数。

18. 如权利要求 15 所述的系统，其特征在于，寻找最近的像素组采用了所述参考图像中的像素组与所述捕获的图像中的选定像素组之间差异的最小平方计算。

19. 如权利要求 15 所述的系统，其特征在于还包括:

过滤器，用于在完成所述其它步骤之前过滤所述捕获的图像以及参考图像。

20. 如权利要求 19 所述的系统，其特征在于，所述过滤包括基

于相邻像素的平均值来确定像素亮度、将所述像素的亮度值与阈值比较、如果所述亮度小于所述阈值则将所述像素的过滤值设定为零以及如果所述亮度等于或大于所述阈值则将所述像素的过滤值设定为 1。

21. 如权利要求 20 所述的系统，其特征在于还包括:

归一化模块，用于通过将已知特征定位在预定位置上来归一化所述捕获的图像和所述参考图像。

22. 如权利要求 15 所述的系统，其特征在于，捕获所述图像是由具有基于视觉的操作系统的视觉处理系统实现的。

23. 如权利要求 15 所述的系统，其特征在于还包括:

比较模块，用于将所述捕获的图像与一个以上参考图像进行比较。

24. 如权利要求 15 所述的系统，其特征在于，比较所述捕获的图像的像素组和所述参考图像的像素组的相对位置包括:

定义所述捕获的图像的第一像素组与所述捕获的图像的第二像素组之间的第一向量;

定义所述参考图像的第一像素组与所述参考图像的第二像素组之间的第二向量; 以及

计算所述第一向量与所述第二向量之间差异的绝对值。

25. 如权利要求 15 所述的系统，其特征在于:

所述计算器比较随机匹配概率与阈值概率，以推断所述图像匹配; 以及还包括:

参考图像检索模块，其用于针对不同的参考图像重复所述步骤，其中，在得到与其中一个所述参考图像的匹配之前，所述系统允许不限数量的不匹配。

26. 如权利要求 15 所述的系统，其特征在于还包括:

图像存储模块，其用于当确定了匹配时将所述捕获的图像添加到参考图像的数据库中，以便于以后的匹配。

27. 一种分析捕获的图像的方法，所述方法包括以下步骤:

识别多个捕获的图像像素组;

识别与所述多个捕获的图像像素组相对应的多个参考图像象素组; 以及

不参考所述捕获的图像的任何已知特征，将所述捕获的图像像素组的相对位置与所述参考图像像素组的相对位置进行比较，以确定所述捕获的图像是否匹配参考图像。

28. 如权利要求 27 所述的方法，其特征在于，所述捕获的图像象素组对应于皮肤块。

29. 如权利要求 28 所述的方法，其特征在于，所述皮肤块是人体上的皮肤块。

30. 如权利要求 27 所述的方法，其特征在于，所述已知特征包括面部特征。

31. 如权利要求 27 所述的方法，其特征在于，识别所述多个参考图像像素组包括识别最佳匹配所述多个捕获的图像像素组的多个参考图像像素组。

32. 如权利要求 27 所述的方法，其特征在于还包括确定所述位置随机出现的概率。

33. 如权利要求 27 所述的方法，其特征在于，所述比较步骤包括:

将所述捕获的图像像素组和所述参考图像像素组分成 n × n 栅格，在参考图像栅格中为捕获的图像栅格中的块寻找最适合的块，并且捕获的图像栅格的块的阈值数在参考图像栅格中具有最适合的匹配时确定匹配。

34. 一种识别捕获的图像的方法，所述方法包括以下步骤:

识别与所述捕获的图像相关联的多个捕获的图像像素组;

识别与所述多个捕获的图像像素组相对应的多个参考图像象素组; 以及

不参考所述捕获的图像的任何已知特征，将所述捕获的图像像素组的位置与所述参考图像像素组的位置进行比较，以确定所述捕获的图像是否匹配参考图像。

35. 如权利要求 34 所述的方法，其特征在于，所述图像是皮肤块。

36. 如权利要求 34 所述的方法，其特征在于，识别所述多个参考图像像素组包括选择对于捕获的图像像素组是最佳匹配的参考图像像素组，并且对于至少一个其它参考图像像素组重复所述选择步骤。

37. 如权利要求 36 所述的方法，其特征在于，比较位置包括:

确定所述参考图像像素组和其相邻之一，及所述捕获的像素图像组和其相邻之一之间的位置关系；以及

比较所述位置关系以确定要识别的图像是否匹配参考图像。

38. 一种图像分析器，包括:

被配置成提供捕获的图像的接口，所述捕获的图像包括多个捕获的图像像素组，以识别人的皮肤的块；和

连接到所述接口的处理器，所述处理器被配置成将选择的捕获的图像像素组的位置与参考图像的选择的参考图像像素组的位置进行比较，所述选择的参考图像像素组对应于所述捕获的图像像素组，其中将选择的捕获的图像像素组的相对位置与选择的参考图像像素组的相对位置进行比较，以估计所捕获的图像和所述参考图像之间的相似性。

39. 如权利要求 38 所述的图像分析器，其特征在于，所述选择的参考图像像素组对应于所述选择的捕获的图像像素组的最佳匹配。

基于视觉的操作方法和系统

技术领域

本发明涉及一种操作方法和系统，特别是涉及一种基于视觉的操作方法和系统。

背景技术

从诸如服务器、个人计算机、膝上型计算机和大型计算机的多功能设备，到诸如设置在各种设备、工具和其它装置中的专用集成电路和处理器的专用装置，计算机系统已变得几乎无处不在。通常计算机是从包括键盘、鼠标、麦克风、摄像机、传感器之类的一组公用接口来进行输入的。然而，虽然对于专用计算机而言，已知道各种各样的输入设备，但是如今用于通用计算机的处理系统却主要集中于两种输入类型：来自键盘、鼠标、触摸屏等的基于字符和基于光标的输入，以及来自麦克风的基于语音的输入（对于语音识别而言）。尽管对于那些计算机存在许多应用，然而在某些情形下它可能难以为用户提供所必需的输入。例如：一个小孩或者老人可能无法足够熟练地提供正确的键盘或语音命令，但是他们能做出可辨别出意义的手势，如求助手势。在另一些情形下，也许没有用户，因此无法得到基于语音或基于字符的指令。不过，给定环境的变化可提供足够的视觉数据，从而触发对相关动作的需求（例如，屋内烟雾可触发警报需求）。因此，需要可接收基于图像或者视觉的输入并对其进行操作的通用处理系统，其中这些输入或者单独或者与常规输入结合起来进行输入。

其中可利用视觉输入而受益的一个领域是面部识别。面部识别技术在如为安全目的处理图像以确定一个面部是否与一个或多个参照考面部匹配方面是众所周知的。这种技术可用于如确定是否准许

用户进入住宅、办公室或类似的环境。目前的面部识别方法，即通常涉及多个图像之间面部特征（feature）的比较和评估匹配度的计算的方法已遇到了一些问题。其中一个问题是这种系统有可能产生假肯定匹配。例如，假肯定结果意味着可准许未经授权的用户进入住宅。产生假肯定匹配的倾向意味着对试图要被识别的用户通常仅做了有限次数的识别尝试。然而，系统也可能产生假否定；因此，防止假肯定匹配所需要的限制趋向于增加假否定数量的增加，这导致在许多情况下会拒绝合法用户的访问。因此，需要一种系统，它可以将假肯定匹配的情况限制到任意低的水平上，以使得想要被识别的用户能尝试进行他/她所希望的很多次识别，而不用担心会准许未经授权的用户进入的情况。

发明内容

在此提供了一种基于视觉的处理系统，它可从环境中得到图像、处理图像（可选地是，不需要另外的用户输入或者交互作用）以及根据图像内容采取动作。一些实施例包括为了安全进入如住宅或者办公室等环境而进行的面部图像处理。其它一些实施例包括利用图像监控环境，例如为了安全目的。本文公开了各种各样实施例，每个实施例都直接利用处理图像的能力，而不需要基于键盘或鼠标的用户输入的干预。

本文还提供了用于确定面部图像是否与参考图像匹配的方法和系统，其中可使假肯定的概率为任意小的值。因此，这种方法和系统允许希望要被识别的用户进行想要的多次尝试直到出现匹配为止，而不用担心会准许未经授权的用户访问。

在一些实施例中，本文所提供的是用于确定已获取的图像是否与参考图像匹配的方法和系统。这些方法和系统提供了：获取数字图像，以便与参考图像比较；将所述获取的图像分成像素组；在获取的图像中识别一组像素；获得参考图像；为像素组检索参考图像，

该像素组与识别步骤中在捕获的图像中识别的像素组相类似；在参考图像中选择与获取的图像中的像素组最佳匹配的像素组；对至少一个其它的像素组重复上述识别、检索和选择步骤；比较参考图像中所选择的像素组和获取的图像中的像素组的相对位置；以及确定相对位置随机出现的概率。

在一些实施例中，这些方法和系统把随机匹配的概率与阈值概率进行比较，以推断图像匹配。阈值可以被设置成使假肯定的概率为任意低。

在一些实施例中，诸如通过根据邻近像素值的平均值来确定像素的亮度，对图像进行过滤；将像素的亮度值与阈值进行比较；如果亮度小于阈值，则把过滤的像素值设置为 0；如果亮度大于或等于阈值值，则把过滤的像素值设置为 1。在一些实施例中，通过把诸如眼睛的已知特征定位到已知坐标上，从而将图像归一化。

在一些实施例中，用于确定获取的图像的像素组与参考图像的像素组之间匹配的方法和系统包括：定义已获取图像的第一像素组和已获取图像的第二像素组之间的第一向量；定义参考图像的第一像素组和参考图像的第二像素组之间的第二向量；以及计算第一向量与第二向量间差异的绝对值。可对所期望的多个像素组进行此计算。

在一些实施例中，本文所说明的方法和系统提供了：将随机匹配概率与阈值概率比较，用于断定图像匹配；对不同的参考图像重复上述比较步骤；以及在得到与其中一个参考图像匹配之前允许不限数量的不匹配。一旦确定了匹配，可将已获取的图像添加到参考图像的数据库中，以便进一步匹配。

在一些实施例中，本文所说明的方法和系统包括基于处理器的系统，它具有基于图像的操作系统。该系统可包括设置在环境中的摄像机，以及与摄像机数据连接的基于计算机的系统，基于计算机的操作系统具有一种操作系统，它可响应于由摄像机所获取的图像数据来操作基于计算机的系统。该操作系统可只根据图像数据来操作

基于计算机的系统。此系统可与另一个系统设在一起，后者能响应于所述操作系统所采取的动作而接收来自基于计算机的系统的指令。其它系统可以是例如安全系统、报警系统、通信系统、自动柜员系统、银行系统、保险箱、另一摄像系统、扬声器系统、麦克风、计算机、服务器、膝上型计算机、掌上计算机、蓝牙使能设备、娱乐系统、电视机、录音机、设备、工具、汽车系统、运输系统、车辆系统、传感器、发射器、发射机、收发器、天线、发送机应答器、游戏系统、计算机网络、家庭网络、局域网、广域网、因特网、万维网、卫星系统、电缆系统、电信系统、调制解调器、电话或蜂窝电话。

在一些实施例中，操作系统可识别图像中的特性，并根据特性采取动作。该特性可能是例如匹配的面部、匹配的代码、动作、生物计量（biometric）、非匹配元素、环境中的结构、面部表情、环境中项目的存在、环境中项目的不存在、项目的运动、图像中新项目的出现、烟、火、水、泄漏、环境的破坏、人的动作、宠物的动作、小孩的动作、老人的动作、脸、手势、摄像机前面部的位置、图像的变化、图像中面部的检测、讲话、嘴唇的动作、手指的动作、手的动作、手臂的动作、腿的动作、身体的动作、头部的动作、颈部的动作、肩部的动作或者步态。在一些实施例中，特性（characteristic）是匹配面部，而动作便是打开安全系统。

本文说明的方法和系统可设在许多环境中，比如机场、飞机、运输聚集场所、公共汽车、公交车站、火车、火车站、出租车聚集场所、汽车、卡车、货车、车间、聚集场所、入场口、运动场、音乐会、露天大型运动场、运动聚集场所、音乐会聚集地、博物馆、商店、住宅、水池、体育馆、健身俱乐部、高尔夫俱乐部、网球俱乐部、俱乐部、停车场、计算机、膝上型计算机、电子商务环境、ATM、存储区域、保险箱、银行、或者办公室。

在系统为允许采取动作而对面部进行匹配的情况下，系统还可要

求更进一步的确认，比如提供钥匙、输入代码、插入卡片、识别声音、识别指纹和识别其它的生物计量。

这里所公开的方法和系统还可包括：将摄像机设置在环境中；捕捉环境图像，所述图像包括环境的事件图像；提供基于视觉的操作系统用于处理图像；处理图像以识别出事件的特性；以及根据所述特性采取动作。

附图说明

图 1 是基于视觉系统的系统部件的高级示意图。

图 2 是视觉处理系统的部件的示意图。

图 3 是图 2 中基于视觉的处理系统的附加部件的示意图。

图 4 是基于视觉的处理系统的备选实施例的示意图。

图 5 是应用了基于视觉的操作系统的视觉处理系统的处理过程的高级流程图。

图 6 是显示基于视觉的处理系统的附加处理过程的流程图。

图 7 说明了可使用基于视觉的处理系统的住宅环境。

图 8 说明了可使用基于视觉的处理系统的运输环境。

图 9 说明了可使用基于视觉的处理系统的车辆环境。

图 10 说明了可使用基于视觉的处理系统的建筑物环境。

图 11 说明了可使用基于视觉的处理系统的计算机环境。

图 12 说明了可使用基于视觉的处理系统的安全存储环境。

图 13 说明了可使用基于视觉的处理系统的聚集地环境。

图 14 是说明采用基于图像的处理系统来保障安全进入住宅的高级步骤的流程图。

图 15 是说明将基于图像的处理系统用于处理手势图像的高级步骤的流程图。

图 16 是说明采用基于图像的处理系统来监控环境的步骤的流程图。

图 17 是说明将基于图像的处理系统用于产品识别的步骤的流程图。

图 18 是说明采用基于图像的处理系统将面部图像与参考图像进行匹配的步骤的流程图。

图 19 是说明图 18 中面部匹配处理的图像过滤处理步骤的流程图。

图 20 说明在应用图 19 中的图像过滤处理之前的面部图像。

图 21 说明应用了图 19 中的图像过滤处理之后的面部图像。

图 22 是说明将面部图像与参考图像进行比较的步骤的流程图。

图 23 说明如何得到面部图像匹配参考图像的确认的步骤。

具体实施方式

参照图 1，提供了一种能够设置在各种环境中的系统 100。系统 100 可包括几个不同的单元，包括摄像机 102 或类似的图像捕获设备和视觉处理系统 104。系统 100 可选地还可包括其它单元件，例如数据存储设备 108 和另一基于计算机的装置 110。

摄像机 102 可以是任何能够捕捉图像数据的装置，例如数码照像机、胶片摄影机、视频摄像机、静止图像摄像机、电影摄影机、beta 录像机、手持摄像机、固定摄像机、感应移动摄像机等等。摄像机 102 可以捕获环境中的图像，并将这些图像传送到视觉处理系统 104。在一些实施例中，图像可作为以包括像素的图像形式的数字数据来进行传送。在其它的一些实施例中，图像可由摄像机以非数字形式来提取，然后由视觉处理系统 104 转换成数字形式，以便处理。摄像机 102 可以配置准许其操作的接口。该接口可以是供人使用的直接用户接口，例如一系列按钮和拨盘，它们可使用户打开和关闭摄像机、记录图像数据、调整镜头的位置、改变镜头参数设置、放大或缩小、录像等。该接口还可以是由例如计算机的另一系统访问的接口，或者是通过该另一系统访问的接口。在一个优选实施例中，视觉处理系统 104 可访问摄像机 102 的接口并控制摄像机 102。

数据存储设备 108 可以是任何合适的用于存储数据的设备，例如 RAM 或 ROM 存储器、文件、智能媒体卡、软盘、硬盘驱动器、盘、数据库、zip 驱动器、数据仓库、服务器、大型计算机或其它用于存储数字数据的适当设备。数据存储设备 108 可包括一种接口，它允许视觉处理系统或用户使用数据存储设备对数据进行存储、控制以及检索，用于任何常规目的。

以下将更详细的论述视觉处理系统 104。视觉处理系统 104 可以从摄像机 102 得到图像数据，并响应于那些图像而采取适当的动作。在一些实施例中，视觉处理系统 104 也可以与数据存储设备 108 配合起来对数据进行存储、控制或检索。在其它一些实施例中，视觉处理系统 104 还可以与其它装置 110 或摄像机 102 配合工作。在某些实施例中，视觉处理系统 104 可向其它装置 110 或摄像机 102 发送控制信号，例如用于激活或定位装置 110 或摄像机 102 的控制信号。在其它实施例中，其它装置 110 或摄像机 102 可向视觉处理系统 104 发送信号，进行可能的交互，或形成传感器反馈回路，这些系统基于环境中的事件或状况、或基于用户与一个或多个系统的交互而相互作用。

在一些实施例中，通信设备 114 可将摄像机 102 与视觉处理系统 104 连接起来。在另一些实施例中，摄像机 102 与视觉处理系统 104 可以集成在一个装置中。通信设备 114 可以是任何用于传输数据的合适设备，例如电缆、电线、网络、无线通信设备、蓝牙设备、802.11 设备、红外、激光、光纤、无线电、电磁、声音或其它通信设备。

另一装置 110 可以是任何能够与视觉处理系统 104 通信的其它装置，例如经由第二通信设备 112 进行通信，第二通信设备可以是与上述通信设备 114 相连的任何所述类型的设备。另一装置 110 可以从一组不同的可能装置中选择，这组装置包括但不限于报警系统、声音系统、传感器、娱乐系统、视频显示系统、安全系统、锁、门、记录系统、测量装置、医疗装置、医药管理系统、器具、烤箱、洗

衣机、脱水机、炉、洗碗机、电冰箱、冷却箱、个人计算机、膝上型计算机、PDA、掌上计算机、服务器、大型计算机、电视机、客户计算机、DVD播放机、立体声系统、VCR、光盘播放机、个人电视录制机、电话以及可视电话。在一些实施例中，视觉处理系统104可以与这些基于处理器的装置或任何其它基于处理器的装置的任何一个集成或者设置在其上。

参照图2，示意图200显示了视觉处理系统104的部件的实施例。视觉处理系统104可包括各种单元，例如处理器202、基于视觉的操作系统204、通信设备208、数据处理设备210和图像处理模块212。

处理器202可以是任何实现处理功能的常规设备，例如微处理器、芯片、集成电路、专用集成电路、板、电路、微控制器、软件、固件或者上述的组合。在一个实施例中，处理器202是基于奔腾的处理器，如用于操作计算机的那些处理器。

下面将进一步详细地论述基于视觉的操作系统204。与主要响应于由键盘、鼠标、时钟引起的事件或类似事件的常规操作系统不同，基于视觉的操作系统配置成可接收图像形式的输入，这些输入替代了或附加可用作常规操作系统的输入的其它事件。因此，基于视觉的操作系统配备有用于处理经数字化为像素的图像并响应于图像的内容采取动作的设备。

通信设备208可以是任何使视觉处理系统104能够与其它系统或装置进行通信或交互的设备，其中其它系统或装置在视觉处理系统104的外部。因此，它可以包括硬件(例如调制解调器、DSL调制解调器、连接器、总线、端口、串口、USB口、网卡等)、软件(通信软件、网络软件等)、固件或其组合。

数据处理设备210可包括例如RAM、ROM、硬盘、存储器卡、智能媒体卡或其它类似的数据处理设备的硬件单元，以及软件单元，例如数据库软件或用于处理任何与数据相关的任务的其它软件，操作系统204可能需要这些软件来与数据存储设备108交互。

图像处理模块 212 可包括用于处理图像的硬件、软件、固件或它们的组合，包括用于在基于视觉的操作系统 204 的控制下执行各种算法和子处理的设备，以便对用作图像处理模块 212 的输入的数字图像进行存储、控制、检索以及采取其它的动作或响应。图像处理模块 212 把图像作为各种信号中的任意一种信号的输入和输出，这些信号包括：针对基于视觉的操作系统的指令；用于存储、控制或检索数据的指令；消息或通信设备 112、114 的其它通信；图像；文本；声音或其它信号。下文将进一步论述优选实施例中的图像处理模块 212 的功能。

参照图 3，在示意图 300 中显示了一个基于视觉的操作系统 204 的实施例的更多细节。在此实施例中，基于视觉的操作系统 204 用作视觉处理系统 104 的主操作系统，因此视觉处理系统 104 从其环境接收的主要输入是图像或其它基于视觉的数据。可选的是，基于视觉的操作系统 204 可控制辅助操作系统，辅助操作系统可以是对基于视觉的操作系统 204 的信号作出响应的常规操作系统 302。该常规系统可以是 Windows、MAC、Unix、Linux 或其它常规操作系统，例如将来可能出现的或开发用于响应于视觉处理系统 104 中的事件或条件而采取动作的操作系统。因此，例如基于视觉的操作系统可以发起由 Windows 操作系统中运行的循环所得到的事件，从而控制视觉处理系统 104 的其它方面，或者内部或外部地发送信号。

参照图 4，如示意图 400 中所述，在另一实施例中，基于视觉的操作系统 204 和常规操作系统 302 的作用可相对于图 3 的配置而对调。在图 4 的实施例中，常规操作系统 302 控制基于视觉的操作系统，用作子系统。在此系统中，常规操作系统可识别包括图像或其它基于视觉的数据的某些输入或事件，并且可以脱离基于视觉的操作系统 204 处理那些输入。常规操作系统 302 可响应于基于视觉的操作系统的输出，这些输出的形式为适合的事件或信号，例如 Windows 事件。因此，常规操作系统 302 在基于视觉的操作系统的设备的协

助下可以控制视觉处理系统 104，以便处理作为输入的图像。

参照图 5，流程图 500 显示了采用基于视觉的操作系统 204 的视觉处理系统 104 的高级处理过程。在步骤 502 上，视觉处理系统 104 捕获图像或基于图像的数据。在步骤 504 上，基于视觉的操作系统 204 或图像处理模块 212 将图像数据转换为信号，以便通知与输入系统的图像相关联的事件。在步骤 508 上，基于视觉的操作系统 204 根据前述步骤中与图像相关联的事件采取动作。

参照图 6，流程图 600 显示了可由这里所公开的系统 100 来实现的其它处理过程。在步骤 602 上，系统 100 可以在环境中配置用于捕获图像的设备，例如摄像机 102。在步骤 604 上，在此环境中可发生图像事件。此事件可能是移动，或者可能是具有特定特性的图像的出现，例如为面部或身体部位或其它有意义的图像。该事件也可以是图像中存在的任一种符号，例如字母、数字、单字、记号或其它符号。该事件还可以是不存在图像的某部分，例如图像中正常出现的项目不存在，或不存在图像中所期望的移动。在步骤 604 中可被识别出的图像事件没有任何限制，并且以下将进一步论述在优选实施例中处理的某些图像事件。一旦在步骤 604 上发生图像事件，那么在步骤 608 便可由摄像机 102 捕获该图像。

接着，在步骤 610 上，该图像可由图像处理模块 212 在基于视觉的操作系统 204 的控制下进行处理。在步骤 612 上，基于视觉的操作系统 204 可输出如由处理步骤 610 所确定的反映图像特性的信号。在步骤 612 中的输出信号可采取任何形式，例如通过常规操作系统 302 处理的事件、控制另一装置 110 的信号、控制摄像机 102 的信号、通过图像处理模块 212 作进一步处理的另一图像或其它反映步骤 608 所捕获的图像的特性的信号或图像。在下文中还会说明处理步骤 610 的某些优选实施例。在步骤 614 上，系统 100 可根据步骤 612 中所建立的信号的内容采取动作。

可在步骤 604 处理的图像事件可包括：将面部定位在摄像机 102

前；检测移动；改变图像内容；检测诸如窗户的特定位置中的图像上的面部；检测人讲话的图像、身体部位的图像(例如唇、手或腿)、用头、手、手指或身体的其它部位作出姿势的图像；面部特征；符号；字母；单字；数字；记号或其它图像，这些图像可能具有对于使用视觉处理系统 104 的目的而言有意义的联系。

在处理步骤 610 上，可对各种各样的图像特性进行分析。例如处理步骤 610 可以确定图像中的对象是什么，例如确定是面部还是人、身体部位、符号、记号或其它特征。同样，处理步骤 610 可以将图像中的对象与另一图像的另一对象进行匹配，例如将面部图像与参考面部的图像进行匹配。处理步骤 610 也可以把诸如代码的另一对象与参考图像中的代码进行匹配。可以在代码、姿势、面部、身体部位、生物计量测量、移动、记号、符号或其它用于各种目的的特征之间进行匹配，在步骤 612 中，输出信号反映是否出现匹配。处理步骤还可以处理环境结构组织的特性，例如在期望的地方某项目(如在一个房间中由摄像机 102 监控的贵重项目)的存在或不存在，或例如窗户、屋顶或门的项目的物理状况，以确保它还未被损坏。处理步骤 610 还可以处理面部的特性，例如由特定的面部移动或位置所表现出的感情。处理步骤 610 也可以处理在环境中是否正产生移动，并输出反映是否正出现任一移动或特定移动的信号，例如用于监控安防环境中的移动，或监控医疗环境中病人的移动以保证偶尔活动的病人的安全。处理步骤 610 还可处理图像以确定在环境中是否已出现了任何新项目，并可分析这一新项目以确定其性质。在处理步骤 610 上，系统 100 可识别特定的环境特征，例如烟、火、湿云雾、水或其它可引起应当从系统发出消息或报警的图像特征。在实施例中，处理步骤 610 可处理图像中的小孩、宠物或其它实体的图像，并根据移动的性质采取动作，例如小孩接近危险的项目，诸如炉子或未受监控的游泳池。在实施例中，处理步骤 610 可以根据任意上述图像特性或事件的组合或其它图像特性或事件来采取动

作，或根据它们中的一个或多个与其它装置或系统的输入的组合来采取动作，例如出于安全目的，将在键盘上的手工安全代码的输入同面部与参考图像进行的匹配结合起来。处理步骤 610 的这些实施例应当理解为众多不同图像特性中的代表，它们可经处理用于识别和采取进一步动作，所有的这些图像特性均应当理解为包含在本公开中。

在动作步骤 614 上可采取许多不同类型的动作。这些动作的例子包括发送消息或其它信息、打开或关闭装置或系统、启动其它装置的动作、引起视觉处理系统 104、摄像机 102 或其它装置 110 的移动或以任何方式对它们进行控制、允许进入安防环境、开锁、发送“全清”信号以及阻止进入环境。当然动作步骤 614 可以发起任何可由任何其它装置 110 采取的动作，因此动作的类型与性质可能是没有限制的。

存在着许多其中可使用基于视觉的处理系统的潜在环境。参照图 7，住宅 700 便是这样的一种环境。可能存在许多的住宅应用。在一种应用中，摄像机 102 可设置在住宅 700 的门 702 上某处，在这里，诸如住宅的居民的用户 704 可看向摄像机 102 以便面部识别(以下将作更详细的介绍)。摄像机 102 可拥有机载视觉处理系统，或与独立的视觉处理系统 104 相连，用于确定用户的面部图像是否与一个或多个存储在数据存储设备 108 中的参考面部图像匹配。如果存在匹配，则门 702 上的锁 708 便可释放，允许进入。在另一实施例中，住宅 700 内设有游泳池 710，在游泳池中设置了水池摄像机 712，用以监控水池环境。水池摄像机 712 可捕获小孩 714 的图像，并且如果小孩 714 太接近于水池 710 的话，可经由视觉处理系统 104 触发报警器 718。这种摄像机与报警器的组合能用来向父母或别的成人发出告警信息：小孩或宠物接近于危险对象，例如炉子、烤箱、壁炉、木炭炉、工作台等，或接近于易碎的项目，例如瓷器、水晶、花瓶或其它贵重物品。

参照图 8，其它可设置视觉处理系统 104 和摄像机 102 的环境为运输环境，例如航线终端安全环境 800。此环境可包括金属检测器 802，以及物品屏蔽装置 804，在两种情况下，这种装置的各种常规类型之一由航空公司用于屏蔽旅客及其物品。在环境 800 中，摄像机 102 可捕获旅客 808 的图像并将旅客的面部图像与参考图像进行匹配，从而作为安全措施来确认旅客的身份。在其它交通安全环境中可设置类似的系统，例如公共汽车、轨道和轮船终端中的那些系统，这些系统也可以设在交通车辆，例如汽车、公共汽车、卡车、飞机、火车、轮船、小船等上。在一个实施例中，交通环境可以是停车点，带有摄像机 102 和视觉处理器 104 的系统 100 可用于监控车辆图像，以确保它不被移动或损坏。如果在预定的时间段期间车辆图像发生了变化，则系统 100 可发出报警声音或向例如车主或保安发送报警信息。系统 100 还可以配备用于识别车主或工作员的面部的设备，以便人可进入并开动车辆，而不会触发报警器或发出报警信息。系统 100 也可用于监控预留停车位置的使用，以便于如果将车停在此处的人的面部与参考图像不匹配，则向停车设施的工作员发送出现了预留停车位置的未授权使用的消息。

参照图 9，视觉处理系统 104 和摄像机 102 可有利地在其中发挥作用的其它环境是车辆，例如汽车 900。摄像机 102 可提取司机 902 的面部图像，并把图像与参考数据库的图象匹配，以确认司机 902 是经授权而驾驶此辆车的。参考数据库可存储确认该司机是车主、该司机是许可司机(即该司机没有违章驾驶或约束驾驶之类的限制)、该司机是租用此车辆的人等数据。当确定匹配时，视觉处理系统 104 可采取各种动作例如开门或锁门、允许或阻止引擎的发动或者允许或阻止其它车辆系统的操作。虽然图 9 中显示的是汽车 900，但是其它交通工具也可采用类似的系统，例如船、卡车、小型货车、出租汽车、公共汽车、轮船、飞机、喷气机、小型摩托车、摩托车等。

参照图 10，另一种环境是建筑物 1000，例如办公大楼、车间或

类似的环境。如同住宅的情况，摄像机 102 和视觉处理系统 104 可用于提供对建筑物 1000 外部或内部的门 1002 的安全访问。同样，摄像机 102 和视觉处理器 104 可用于监控一个或多个建筑物 1000 中的项目，例如为防止它们被盗窃或为了监控它们在此建筑物内的位置。例如，视觉处理系统 104 可以基于项目的图像形状、或基于项目上所找到的代码或图像（例如品牌或标志）来确定仓库中库存项目的位置。视觉处理系统 104 随后可与诸如库存控制系统的另一计算机系统交互。在另一实施例中，摄像机 102 和视觉处理系统 104 可用于监控人 1008 的动作，例如出于安全目的以确保此人是有意识的并且未受到伤害，或出于其它目的，例如为确定某动作是否已经出现，该动作可能是采取下一步动作的前提。例如，系统 100 能确定其图像与数据库中的参考图像相匹配的项目已从预处理位置到达作进一步处理的位置的时间，然后向用户 1008 发送项目已准备好的消息。可以设想出视觉处理系统 104 中的基于视觉的操作系统 204 在车间的许多其它应用，并且这些应用应当理解为包含在本说明书中。

参照图 11，摄像机 102 和视觉处理系统 104 可用在含有诸如个人计算机、膝上型计算机、个人数字助理、掌上计算机等的计算机 1100 的环境中。摄像机 102 可捕获计算机 1100 的环境中的图像，例如用户 1102 的图像。视觉处理系统 104 可以设在摄像机 102、计算机 1100 或在它们两个外部的另一计算机系统内，可处理由摄像机 102 提取的图像，如用户 1102 的图像。例如，视觉处理系统 104 可以将用户 1102 的面部与一组参考图像匹配，以确认用户 1102 的身份。因此，为安全起见，该系统可代替或附加其它安全措施，例如口令。在一个实施例中，例如出于电子商务、游戏、研究或娱乐目的，计算机 1100 可由用户 1102 用来同站点（例如因特网站点）进行交互。在电子商务应用中，用户 1102 可利用视觉处理系统 104 来确认用户的身份，从而确保电子商务交易(如使用信用卡购买商品或在线服务)的安全。

参照图 12，在另一实施例中，系统 100 的摄像机 102 和视觉处理系统 104 可设在提供安全存储（例如现金或其它贵重物品）的环境中，这些环境例如为自动柜员机(ATM)1200。系统 100 则可用于在准许一项交易（例如取款、检查帐户余额或转帐）之前验证用户 1202 的身份。同样，系统 100 可用于提供身份验证，以便可以安全存取、从安全的地方取出贵重物品或从锁着的任意大小的存储设备中取出项目。在实施例中，系统 100 可提供身份验证功能和独立的功能，例如监控设备中所保护的项目的图像，以保证它们不被取走、移动或损坏。

参照图 13，包括摄像机 102 和视觉处理系统 104 的系统 100 也可设在包括聚集场所 1300 的环境中，该聚集场所包括门 1302 或用以限制进入聚集场所 1300 的类似设施。聚集场所 1300 可以具有中央计算机系统，或在带有每个系统 100 的门上可以包括计算功能。为了对由摄像机 102 所提取的图像进行匹配，系统 100 可访问参考图像数据库，以确保正试图访问聚集场所 1300 的用户 1304 是已授权用户，例如确认用户 1304 为聚集场所的事件买了票，或用户 1304 是雇员或经授权有资格进入该聚集场所的订约人。可以设想出众多不同的聚集场所，例如体育运动竞赛聚集场所，例如足球、篮球、英式足球、曲棍球、垒球和高尔夫球聚集场所；表演聚集场所，例如电影院、剧场、事件中心、音乐会聚集场所和歌剧院；膳宿聚集场所，例如宾馆、汽车旅馆、娱乐场、酒吧、会议中心和饭店以及许多其它的聚集场所。

参照图 14，流程图 1400 显示了本发明的一个实施例的高级步骤，其中系统 100 用于保证进入诸如住宅的环境的安全。在步骤 1402 上，系统 100 可捕获用户的面部图像(或其它可识别的特性)。接着，在步骤 1404，系统可将此图像与一个或多个存储在数据设备的参考图像进行比较。然后，在步骤 1408，系统可确定该图像是否匹配(如下文所作的更详细的说明)。如果不匹配，则系统可通过返回到图像捕获

步骤 1402 再进行尝试。如果匹配，则在步骤 1410，系统可准许进入环境。

参照图 15，流程图 1500 显示了本发明的一个实施例的步骤，其中视觉处理系统 104 处理来自摄像机 102 的图像，以便可以识别以及按图像中所捕获的姿势动作。在步骤 1502，摄像机 102 捕获可能包括姿势的图像并将其转递到视觉处理系统 104。在步骤 1504 上，视觉处理系统 104 的图像处理模块 212 将所捕获的图像与数据库的图像姿势进行比较，从而确定所捕获的图像是否包含与所存储的姿势匹配的姿势。接下来，在步骤 1508 上，图像处理模块 212 确定是否已出现了匹配。如果不匹配，则处理返回到步骤 1502 继续捕获图像。如果在步骤 1508 上出现了匹配，则在步骤 1510 上，系统通过参考已存储的规则来确定何种姿势已匹配以及何种动作合适，上述已存储的规则把每个姿势或一系列姿势与关联动作相关。接下来，在步骤 1510 上，系统根据所识别的姿势发起动作。在某些情况中，该动作可以是等待下一姿势，以便系统可根据组合姿势以及根据单个姿势来动作。举例来说，系统可监控病人并触发该病人是否情况良好的询问。如果病人作了“竖起大拇指”手势的姿势，则该系统可向护理人发送病人情况良好的消息。同样，该系统可捕获例如摇手的姿势，以指示应该触发报警器或发出报警信息。通过创建一套完整的规则，视觉处理系统 104 有可能发起任何原本由键盘、鼠标或声音输入触发的动作。因此，视觉处理系统 104 通过姿势控制可替代或补充常规计算机操作系统。

参照图 16，流程图 1600 表示采用系统 100 来监控环境以便触发适当动作的高级步骤。在步骤 1602 上，该系统捕获环境图像。接下来，在步骤 1604 上，该系统将所捕获的图像与数据库的参考图像进行比较，以确定是否出现匹配。如果在步骤 1608 上未出现匹配，则处理返回到步骤 1602 进行下一图像的捕获。如果在步骤 1608 出现了匹配，则在步骤 1610，系统可访问多个规则，这些规则用于确定

响应于图像的识别而应当采取的动作。然后，在步骤1612，该系统可基于这些规则发起动作。可被匹配的图像的例子包括显示移动的图像、显示接近于特定项目的移动的图像、具有唯一特征性（如烟、火或水）的图像、显示两个项目相互接近的图像(例如为防止碰撞)的图像、没有移动等等。当这些项目其中之一受到匹配时，这些规则随后便可确定动作。例如，如果检测出烟、火、或水处于异常状态，则向操作员或应急服务机构发送报警或消息。如果两个项目(例如两艘船只)行驶的过于接近，则可用声音向操作员发出报警。如果小孩太靠近水池或炉子，则可发出报警声音并向父母发送消息。如果图像中的项目缺失，则可向保安或其它负责监控该项目的人发出报警信息。因此，通过图像匹配和触发规则，该系统可广泛地用于任何环境的监控。

参照图17，流程图1700显示了在用于产品识别的实施例中采用本发明的高级步骤。在步骤1702上，摄像机102可以用于捕获产品图像，例如在仓库中为控制库存或在零售柜台上为确定价格。然后在步骤1704，可将产品图像与存储在数据设备中的图像进行比较。在步骤1708，系统可确定是否已经出现了匹配。如果没有出现匹配，则该系统可返回到步骤1702并继续尝试寻找匹配。如果出现了匹配，则在步骤1710，该系统可确定从中捕获图像的产品的性质，并咨询针对该产品所存储的用于确定动作的一组规则。例如，此系统可确定项目是一箱特定品牌的谷物，可依据检索价格的规则检索出那箱谷物的价格。接下来，在步骤1712，此系统可基于图像的确定结果发起动作，例如对这箱谷物计价，或存储特定项目在特定位置中的指示。

下面将提供其它有关用于匹配面部以确认系统100的用户身份的系统的细节，例如经由住宅安全系统准许进入住宅的系统。如图7所示，系统100可用在门702上或入口通道上，从而可通过锁708或类似机制来控制进入住宅700。如图14的流程图1400所示，系统

可用于将面部与数据设备中的一个或多个参考图像进行匹配。与图 1 相同，面部匹配系统可具有类似于更通用的视觉处理系统 100 的部件，例如摄像机 102、视觉处理系统 104、数据设备 108 以及可选的其它系统 110，如电子开锁系统。

参照图 18，流程图 1800 公开了面部匹配系统的实施例的步骤。首先，在步骤 1802，该系统例如通过用户将他或她的面部放在摄像机 102 的前面来获得面部图像。然后在步骤 1804，视觉处理系统 104 的图像处理模块 212 过滤初始图像，从而获得已过滤的图像，它更适用于匹配目的。图 19 的流程图 1900 中公开了过滤步骤的其它细节，图 19 与图 18 通过分页连接符“A”相连。此过滤步骤将图像分解成像素矩阵，然后将每个像素的相邻像素的亮度进行平均，再根据给定像素的平均是否超出了二进制阈值而把图像分成若干像素。当在步骤 1804 将图像过滤后，便开始进行匹配处理的其它步骤。在步骤 1808 该系统可以把过滤的图像分成多个像素组或像素块。步骤 1808 可包括可选的归一化图像的预处理步骤，例如把眼睛或面部的其它特征设置在与参考图像的眼睛相同的位置中。按图像的列和排或行，可以将它分成块，例如分成 n 维的方块。这些块的数量是两个因子的乘积，两个因子分别为由 n 所除的列数以及由 n 所除的行数。可把这些方块考虑为其中心在像素(i，j)上，其中 i 是 0 和列数之间 n 的整数倍，j 是 0 和行数之间 n 的整数倍。然后，在步骤 1810，该系统获得来自数据设备 108 的参考图像，该参考图像可以存储在本地或存储在远端主机上。接着，在步骤 1812 上，系统搜索第一参考图像。对于任意多的参考图像，可按顺序以并行的方式重复步骤 1810 和 1812。例如，所存储的可以是显示处于不同条件的用户的不同图像，例如晒黑、未晒黑的用户图像等。接下来在步骤 1814，该系统应用一种或多种算法来确定所捕获的图像与参考图像之间的差异，这些算法将在下文进行详述。结合图 22 公开了其它步骤，图 22 通过分页连接符“B”与流程图 1800 相连。当在步骤 1814 计算出差

异后，便在步骤 1818 确定是否存在匹配。用于评估和确定匹配的步骤在图 23 中有更详细的说明，图 23 通过分页连接符“C”与流程图 1800 相连。一旦已确定出匹配，该系统便可在步骤 1820 发起动作，这些动作可包括允许对设备的访问，并且可选地是，包括把新捕获的图像存储在参考数据库中以便用于以后的匹配。如果没有匹配，则该系统可重复上述步骤。因为匹配阈值可以为任意困难，所以假肯定匹配的概率可以任意低，使得适当允许进行多次甚至不受限制的匹配尝试，而不像许多常规系统那样必须防止大量的尝试，因为大量的尝试会导致允许非法访问的假匹配的概率的增大。

参照图 19，流程图 1900 通过分页连接符“A”与图 18 的流程图 1800 相连，其公开了用于实现图 18 的步骤 1804 上的图像过滤步骤。首先，在步骤 1902，为以数学方式说明图像，可将它视做像素矩阵 p_{ij}。像素一般是三种颜色红、蓝、绿颜色的重叠，因此可把各 $p_{i,j}$ 作为向量空间 R^3 的一个元素。该像素的三个分量表示一种颜色根据基色的分解。为简化起见，可忽略各个分量的离散特性，并认为每种颜色的每个细微差别是允许的。因此，n 行和 p 列的图像可描述为矩阵 $A \in M_{np}(R^3)$。

接下来，在步骤 1904 上可以定义各个像素的亮度 L 为从 R^3 到 R 的线性函数，例如：L: (r,g,b)→0.3r+0.59g+0.11b。接着，在步骤 1908，可以根据周围像素的亮度定义像素平均值。例如，可把 Average 定义为像素 p_{ij} 的相邻函数，并给出全部像素($p_{k,l}$)的亮度的算术平均，其中 $i-r \le k \le i+r$，$j-r \le l \le j+r$，如下所述。

即：$\mathrm{Average}(i, j, r) = \dfrac{\sum_{k=i-r}^{i+r}\sum_{l=j-r}^{j+r} L(p_{k,l})}{(2r+1)^2}$

其它许多函数也可以完成与此相同的任务，例如函数 L 的多种离散卷积，或甚至是非线性的或其它函数。接着，可基于相邻像素的平均对像素进行过滤。因此，在步骤 1910，可定义该过滤，其为根据 Average 值确定的二进制标志：

$$\text{Filter（i,j,r,阈值）} = \begin{cases} 1，\text{如果 Average(i,j,r)-L}(p_{i,j})>\text{阈值} \\ 0，\text{如果 Average(i,j,r)-L}(p_{i,j})\leq\text{阈值} \end{cases}$$

因此，若干变化颜色和亮度的像素可转换成由值为 1 和 0 的像素定义的黑色和白色图像，其中 1 和 0 的值是通过周围像素的亮度是否超过给定的阈值来确定的。在步骤 1910，该系统可输出过滤的图像。这种过滤更加剧了图像中的对比度，以便可以对重要的面部特性（例如疤痕、胎块等）进行匹配。参照图 20 和 21，可以看到在进行例如联系图 19 所公开的过滤处理之前的面部图像(图 20)和之后的面部图像(图 21)。

参照图 22，在流程图 2200 中还公开了在捕获的图像和参考图像之间作比较的步骤，该流程图通过分页符“B”连接到图 18 的流程图 1800。在下文中，规定下标 1 针对与参考图像相关的量，下标 2 针对捕获的图像。参照这两幅图像 I1 和 12，它们已被分成相同的格式列 × 行。这些图像已经是归一化的图像，例如，位于标准模式中的两幅图像的眼睛的位置相同。两者均已例如经过图 19 的流程图 1900 的步骤进行了过滤，成为两幅基本上相同大小的二进制图像。实质上，图像的比较可理解为用于判断两幅图像的两个正方块存在的差异程度的误差函数。在整个下文中，假定一个已选择 n 维方块的格式。在流程图 1800 的步骤 1808 上，已将第一图像分成若干完全分开的像素方块。有列/n*行/n 个这样的方块。每个方块可由以像素(i_1, j_1)为中心的 $B_{i1,j1}$ 设计，其中 i_1 是 0 和列数之间的 n 的整数倍，j_1 是 0 和行数之间的 n 的整数倍。应当注意，这里所说明的方法也可用于其它块状，如圆形、矩形等。这里是为了便于说明才采用方块表示的。

在步骤 2202 中初始化计算，开始，按照某种误差函数计算第一图像的各个方块中哪一个最配合第二图像中的方块。在步骤 2204，针对捕获的图像中的第一像素块计算误差函数，该误差函数由那一

像素块的像素与每幅参考图像的每个可能像素块的像素之间的差异的平方和构成。在步骤 2208 选择作为与那一像素块“最佳”匹配的参考图像的像素块。然后，在步骤 2210，该系统针对近处的像素块存储参考图像的最佳匹配像素块的位置。接着在步骤 2212，该系统确定是否还有其它像素块要分析。如果有，则重复步骤 2204 到 2210，直到参考图像中的每个像素块在已知的位置上已识别为是最佳匹配的为止。当这些步骤完成后，如分页连接符 B 所指示的那样，处理返回到流程图 1800。

在步骤 2208 上确定最佳配合可用多种技术来实现，这些技术包括最小化差异之和、最小平方和其它类似差异计算函数。当处理完成时，可在第二图像中识别出以(i_2，j_2)为中心的块，其最佳对应于第一图像中以 (i_1，j_1) 为中心的块 $B_{i1, j1}$。该块是 $B_{i1, j1}$ 的‘全局最佳配合’。

参照图 23，流程图 2300 说明了可评估摄像机 102 所捕获的第一图像是否匹配从数据存储设备 108 检索到的第二图像的步骤。图 23 的处理可以可在基于视觉的处理器 204 的控制下用图像处理模块 212 来实施。

图 23 所述的匹配处理利用了连续性原理，具体来说利用了要匹配的面部图像的特殊连续性。一般地，当比较同一面部的两幅图像时，如果已经在两者中定位了某特定点(第一图像中的特定点称为 Ml，第二图像中的特定点称为 M2），可预测恰好在 Ml 的右边(第一图像中)的细节 N1 应该对应于恰好在 M2 的右边(第二图像中)的点 N2。因此，期望在两幅图片中对于 Mi 而言 Ni 的相对位置是相同的，或几乎是相同的。(可能认为这两张图片近似地是通过以某种方式平移图像而相互导出的)。如果怀疑两幅图像中的点 Mi 的对应，则如果两图中相邻的 Ni 具有相似的相对位置，那么将会更加相信图像的匹配。对应出错的机会更小，比以前小两倍。

继续此分析，作两种对立的假设，其可具有如下特性。使 H0 作

为假设：有“两个不同人”的两幅图像。使 H1 作为另一假设：有“同一个人”的两幅图像。

可定义(i_1, j_1)的邻居，称为(k_1, l_1)。实际中，用模方‖‖来表示邻居$\|(i_1-k_1, j_1-l_1)\|\leq 1$，可称$(i_2, j_2)$为$(i_1, j_1)$的全局最佳匹配，$(k_2, l_2)$为$(k_1, l_1)$的全局最佳匹配。下面得到基本的概率性假设。“期望”当处理同一个人(H1)的图像(这是连续性的自变数)时(k_2, l_2)为(i2, j2)的“邻居”，在此意义上$\|k_2-i_2+i_1-k_1, l_2-j_2+j_1-l_1\|$接近于 0。相反，“期望”当图像来自两个不同的人(H0)时，(k_2, l_2)应当以相同概率地处于所搜索区域的任何地方。例如，可以设想在（k_1，l_1）与可能的（k_2，l_2）之间的误差标准是一串独立且同一法则的随机值。这些值中的最大值其本身是均匀分布在搜索区域中。

首先在步骤 2302，定义系统在其中寻找最佳配合尺寸 S 的区域，接着，在步骤 2304，操作员定义允差参数 t，并标识出两幅图像中的两个块的相对位置的两个向量 $V_1=(i_1-k_1, j_1-l_1)$，$V_2=(i_2-k_2, j_2-l_2)$。认为当处理同一个人（H1）时,V_1和 V_2接近，相反，当处理不同的人（H0）时，V_1和 V_2独立且为一致法则。

下面将集中于以下事件：$\{\|V_1-V_2\|<t/2\}$，可定义那个事件为“确认”。此事件对应于第二图像中的(i_1, j_1)和(k_1, l_1)全局最佳匹配的‘连续性’事件。等效地，此事件可以描述为其中(k_1, l_1)的全局最佳匹配与(k_1, l_1)的‘局部’最佳匹配一致的事件，局部最佳匹配是仅在$(i_2+k_1-i_1, j_2+l_1-j_1)$周围的小区域中所寻找的。此事件正好对应于 V_2的$(t-1)^2$个可能值。如前所述，此事件在不同人的假设(H0)中概率小，对于 V_2的可能值的总数为 S。在假设(H0)中，根据一致性假设，有：

$P((k_2,l_2)$“确认”$(i_1,j_1)|H0)=P(\|V_1-V_2\|<t/2|H0)=(t-1)^2/S$

可对(i_1, j_1)的其它邻居作类似计算，假设这些中的 I2 中的所有最佳配合均是独立放置的，总遵循随机一致法则。

有条件地针对(i_2, j_2)的位置，发现 p 个中有 k 个(i_1,j_1)的邻居在图像 I2(允差取模)中与图像 I1 中相同的相对位置中存在其最佳匹配的

概率是:

$$P((i_1, j_1)\text{的邻居中的 k 个 ‘确认’}|H0)=C_p^k[(t-1)^2/S]^k[1-(t-1)^2/S]^{p-k}$$

考虑数字的例子。使$(t-1)^2/S=10\%$，使 k=3 和 p=4。有 0.36%的 k 个确认的概率。通过从 4 个里找到 3 个(i_1, j_1)的邻居，在图像是不同的假设下，此事件的概率非常小，即该事件是难以相信的。

前述涉及的仅仅是图像 I1 中的一个块 Bi 及其紧接着的邻居。分析假定邻居的最佳配合可均匀地分布在搜索区域中。如果按照同一思想处理，不是从$B_{i1, j1}$开始而是从其它块（例如$B_{i1',j1'}$），可以首先发现它在图像 I2 中的绝对最佳配合，称为(i_2', j_2')。以类似方式，随后可在中心为(i_2', j_2')的搜索区域中寻找$B_{i1',j1'}$的邻居的局部最佳配合。

紧记搜索的局部表面$(t-1)^2$与图像的总表面 S 相比非常小，这是有益的。这意味着在实际中，在表面$(t-1)^2$中所搜索的某局部最佳配合很少机会成为整个图像的全局最佳配合。

现在可以总体地记下图像 I1 的全部块，它们以称为$(i_{1,1}, j_{1,1})$、$(i_{1,2}, j_{1,2})\ldots(i_{1,N}, j_{1,N})$的位置为中心。对于第一块在 $(i_{1,1}, j_{1,1})$的邻居之间具有 k_1 个‘确认’的可能性，对于第二块具有 k_2 个‘确认’的可能性…直到最后一块具有 k_N 个‘确认’的可能性。这里 ε 代表$(t-1)^2/S$。

$$P((i_{1,1}, j_{1,1})\text{的 } k_1 \text{ 个‘确认’，…}(i_{1,N}, j_{1,N})\text{的 } k_N \text{ 个‘确认’}\|H0)=\prod_{q=1}^{N} C_p^{k_q}(\varepsilon)^{k_q}(1-\varepsilon)^{p-k_q}$$

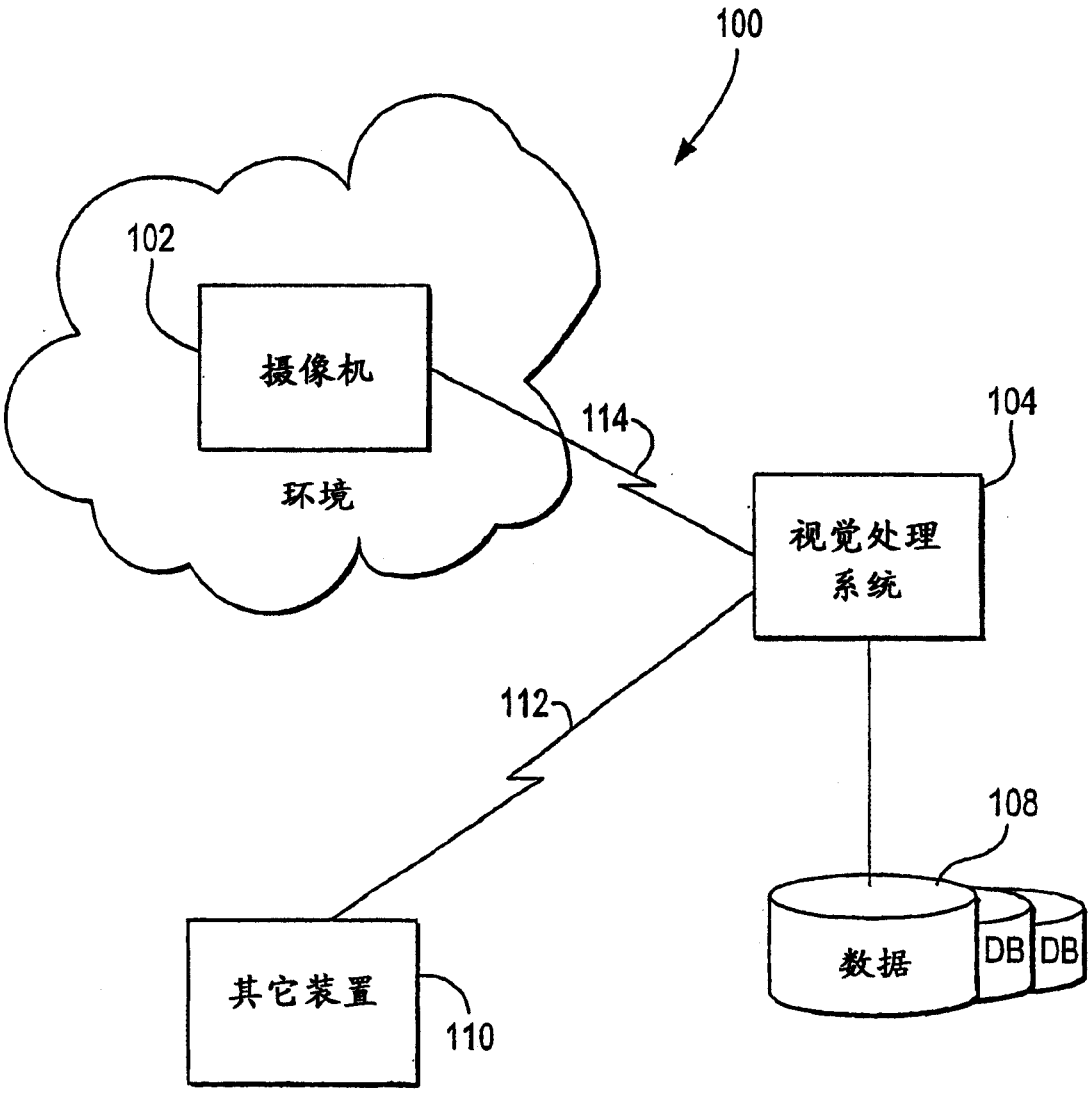
调用按照两幅图像是两个不同人的假设进行的计算。前面的计算给出了全部块中的确认事件的共同规则。这意味着实际上可以准确地评估假肯定匹配的概率，即,使两个不同的人的两幅图像高度相似的概率。实际上，错误识别的概率几乎为零。

利用连续性原理可以完成流程图 2300 用于确定匹配的过程，其中，连续性原理允许创建判别规则（very discriminant rule）以使随机出现的相邻像素块的匹配分开，这与由于真正的相干对应(因为包含像素块的面部是相同的)所产生的匹配相反。在步骤 2306 上，系统计

算各个像素块的最佳配合。在步骤2308，系统然后确定合成统计指示器考虑出现全部‘确认’事件。在步骤2310，如果前述统计指示器大于在步骤2304所定义的阈值，则该系统宣称匹配，或者如果相反的情况则宣称不匹配。我们知道，在H0下，统计指示器大于其阈值的概率是意外的，这与由于真正的匹配所产生的概率相反。因此，通过定义此阈值，对于操作员而言有可能建立任意苛刻的匹配准则，从而把假肯定概率降低到任意低的水平。

虽然这里已公开了某些优选实施例，但是本领域的普通技术人员会认识到其它实施例，这些实施例如仅由权利要求所限制的那样应当理解为包含于本发明中。这里所参考的所有专利、专利申请及其它文档通过引用结合于本文中。

图 1



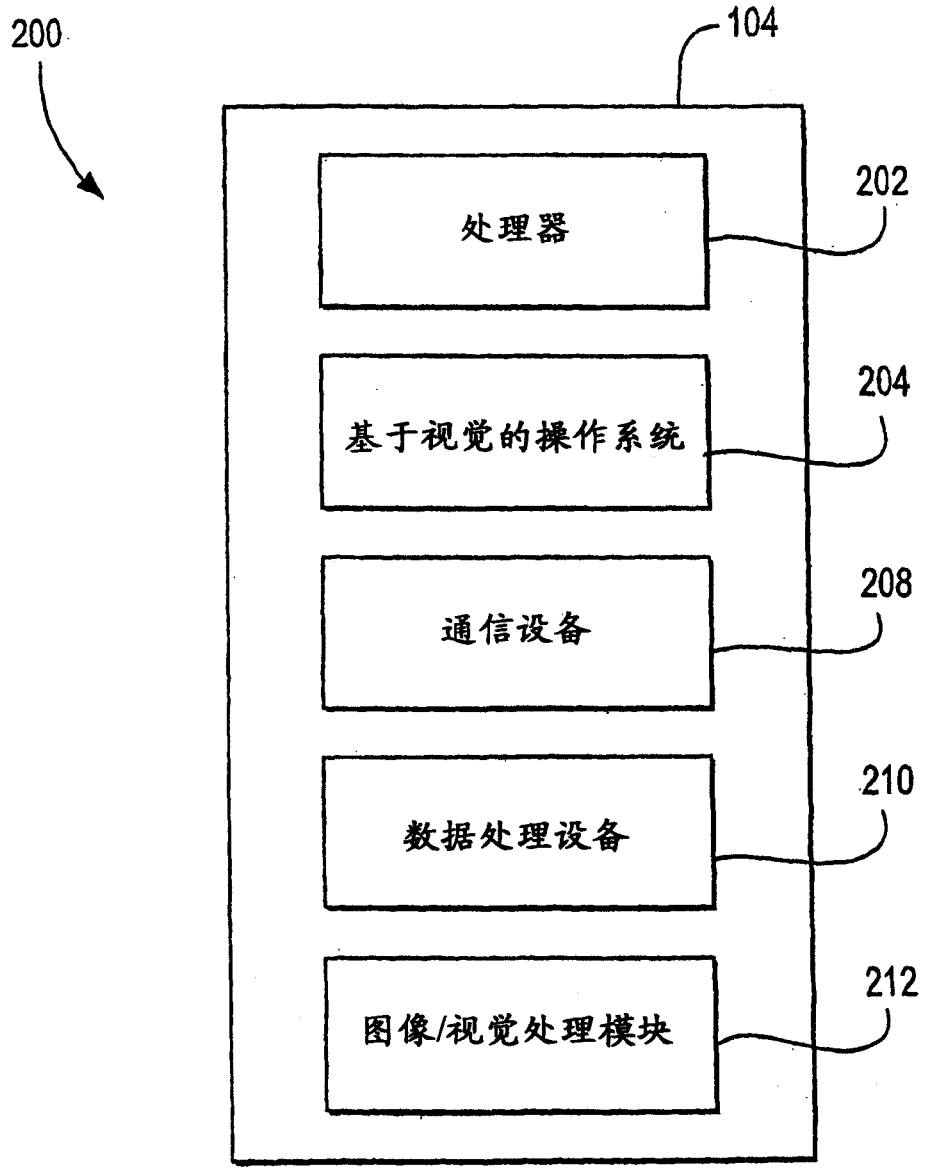

200
104
处理器
202
基于视觉的操作系统
204
通信设备
208
数据处理设备
210
图像/视觉处理模块
212

图 2

图 3

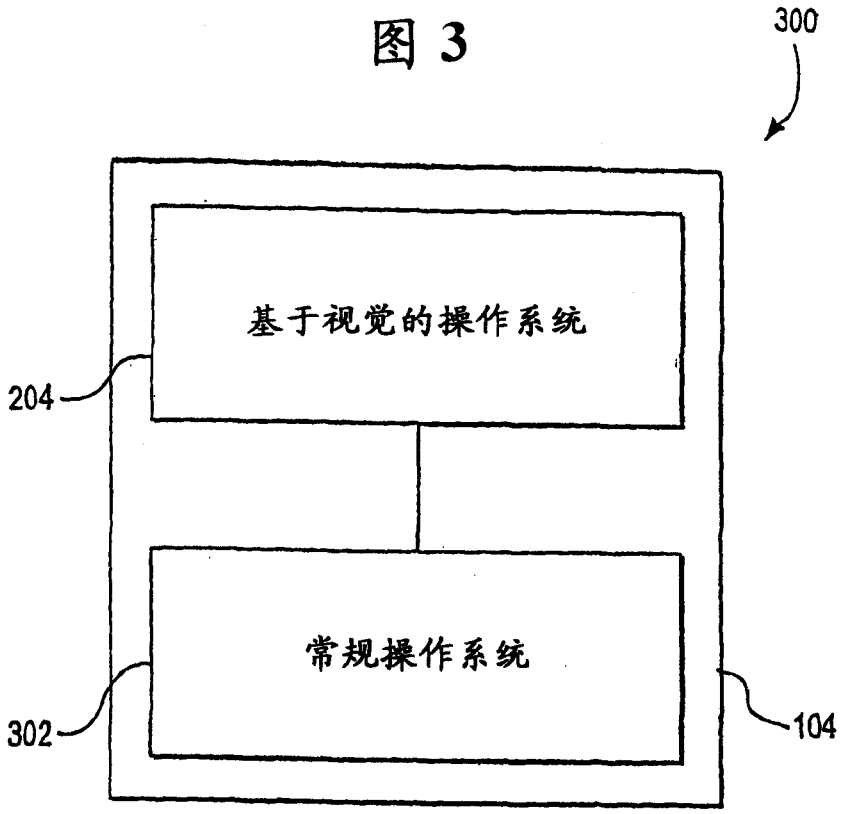
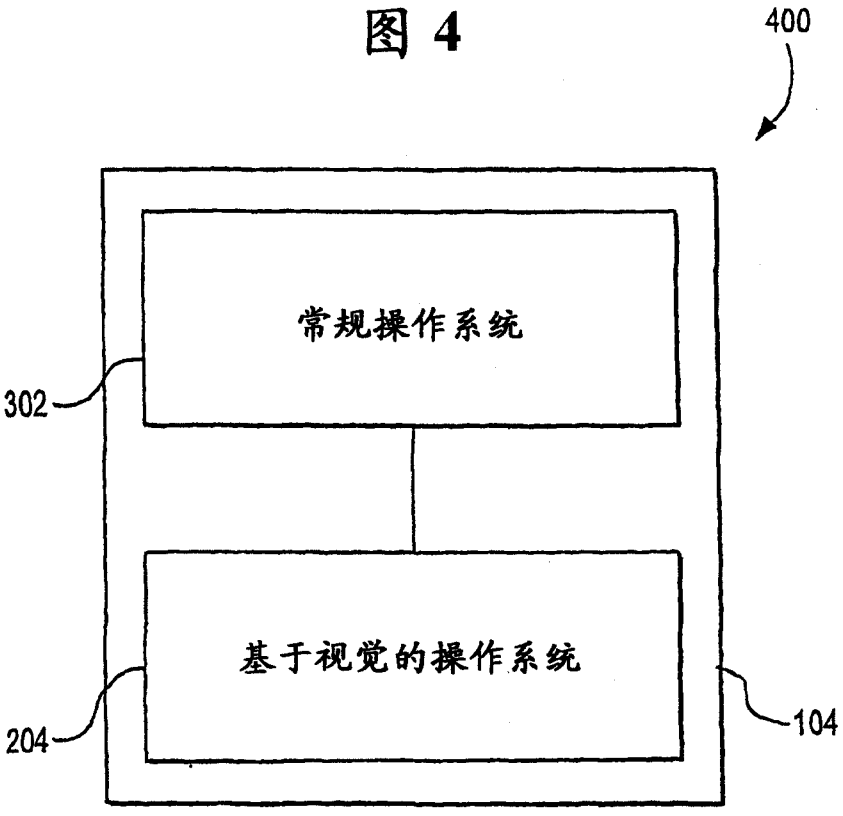

300
基于视觉的操作系统
204
常规操作系统
302
104

图 4


400
常规操作系统
302
基于视觉的操作系统
204
104

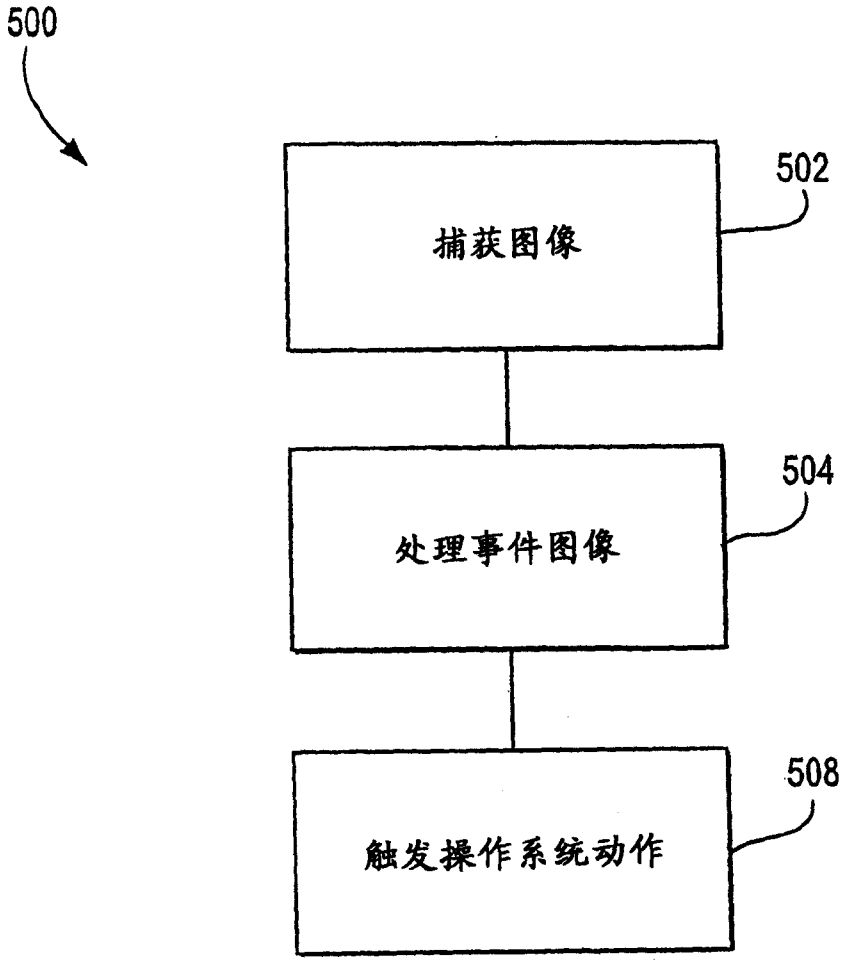


图 5

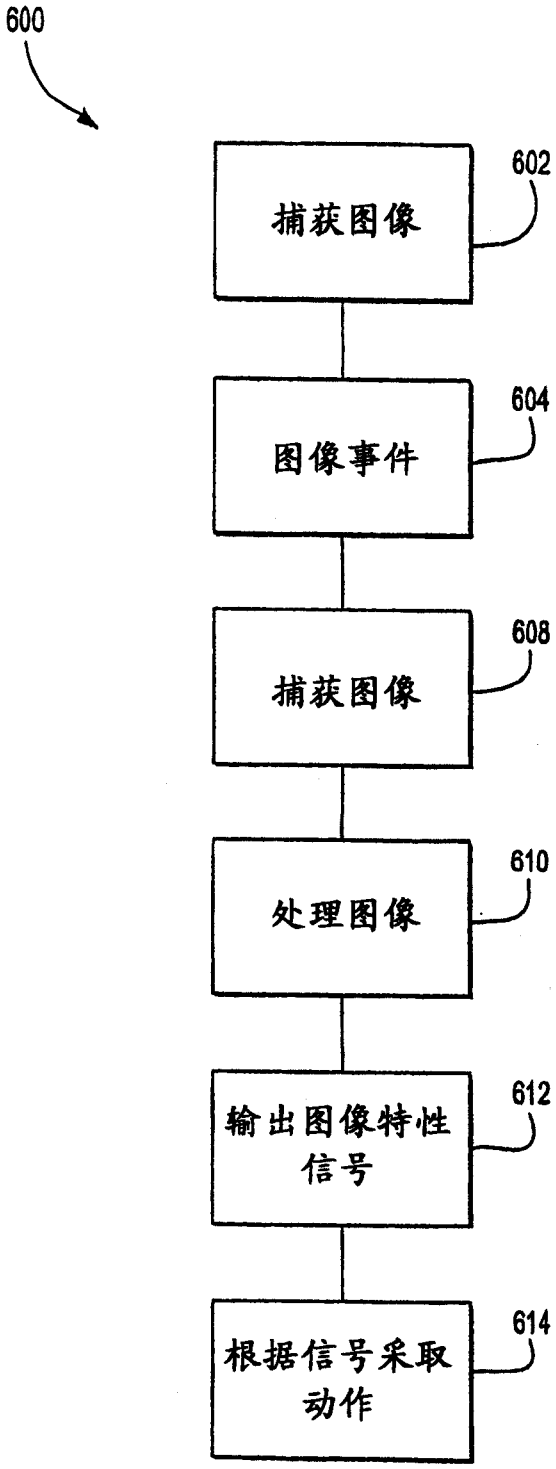

600
捕获图像
602
图像事件
604
捕获图像
608
处理图像
610
输出图像特性信号
612
根据信号采取动作
614

图 6

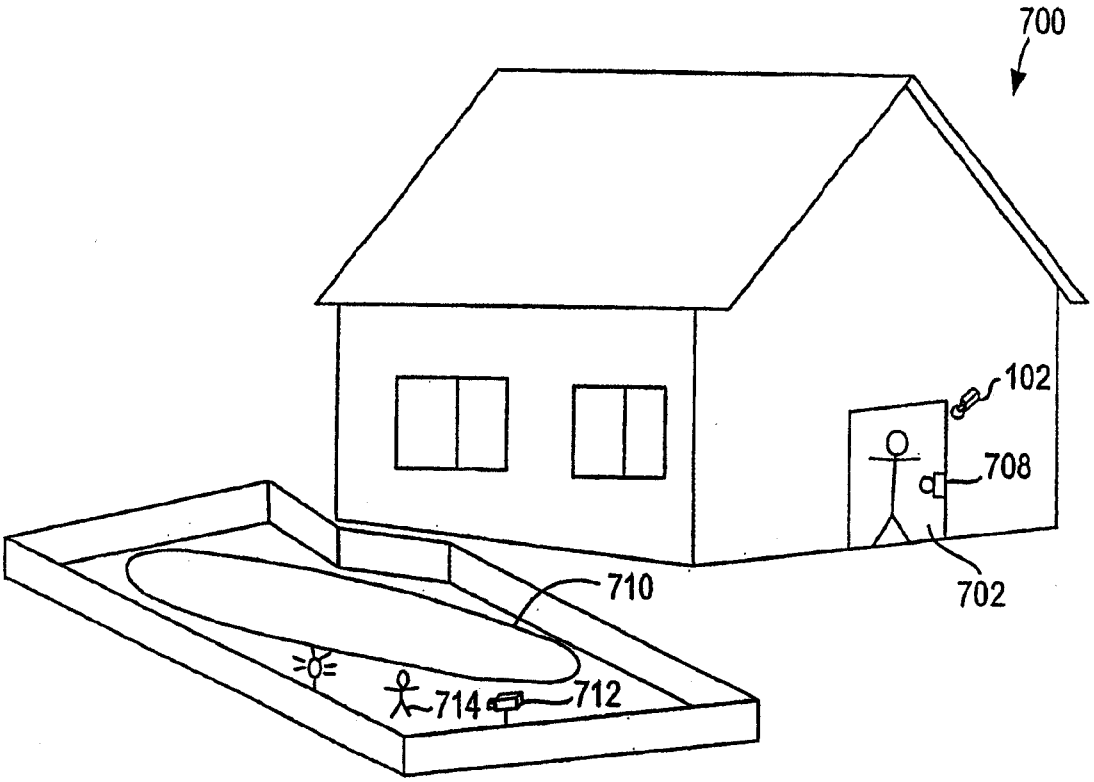


图 7

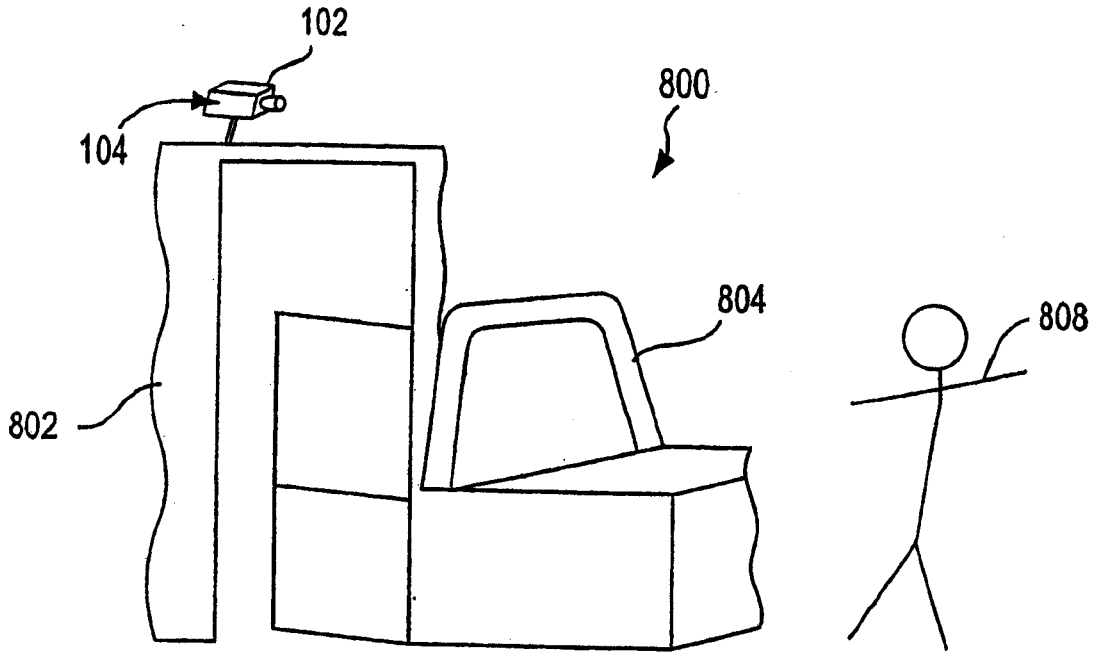


图 8

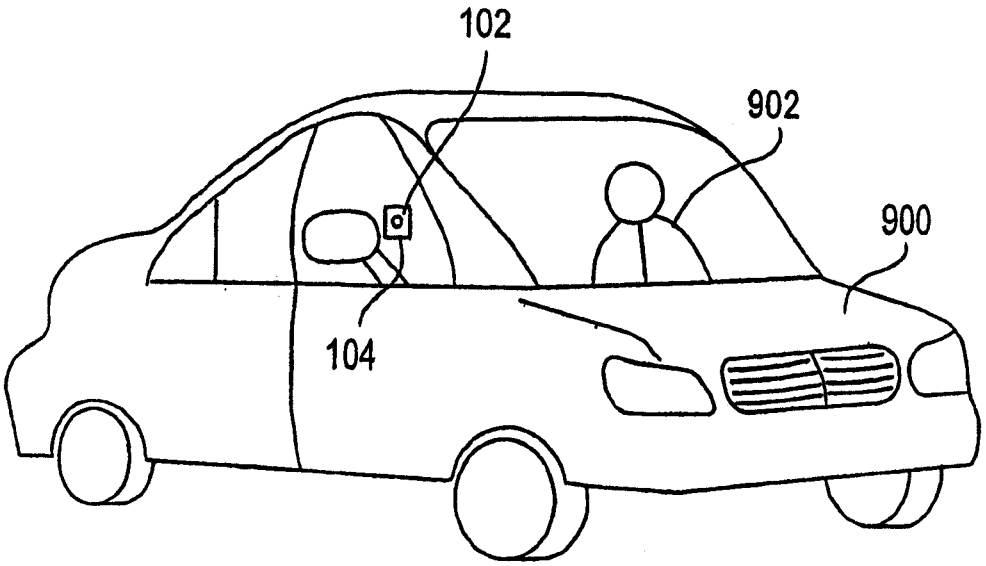


图 9

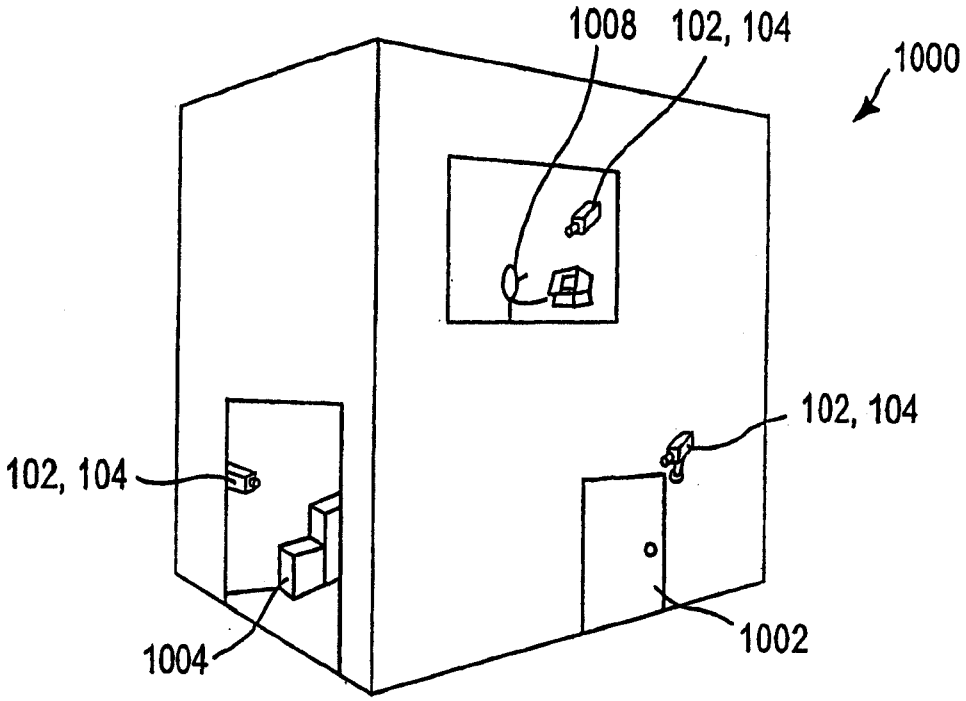


图 10

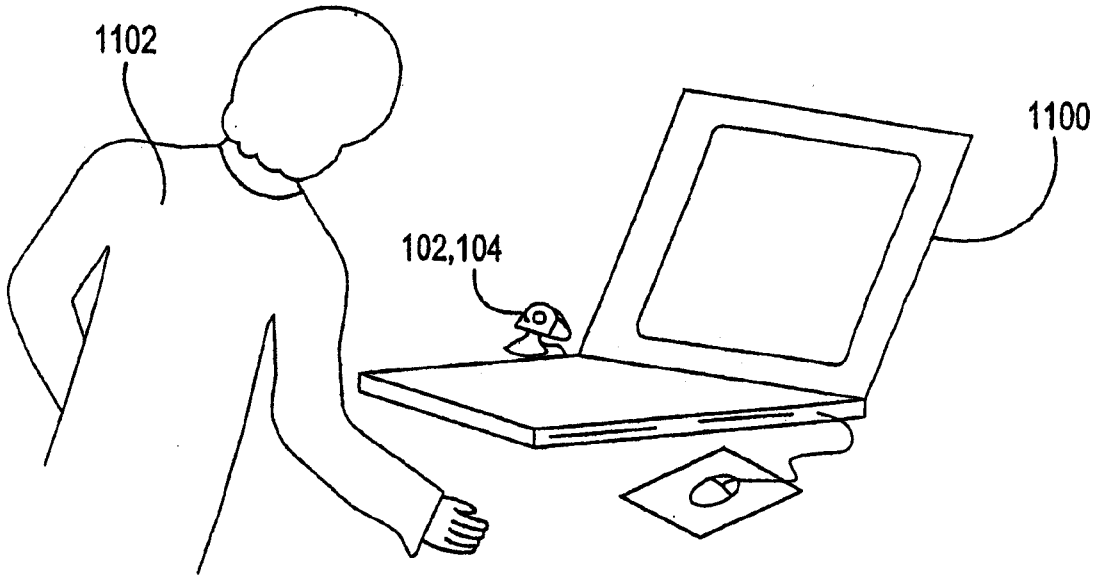


图 11

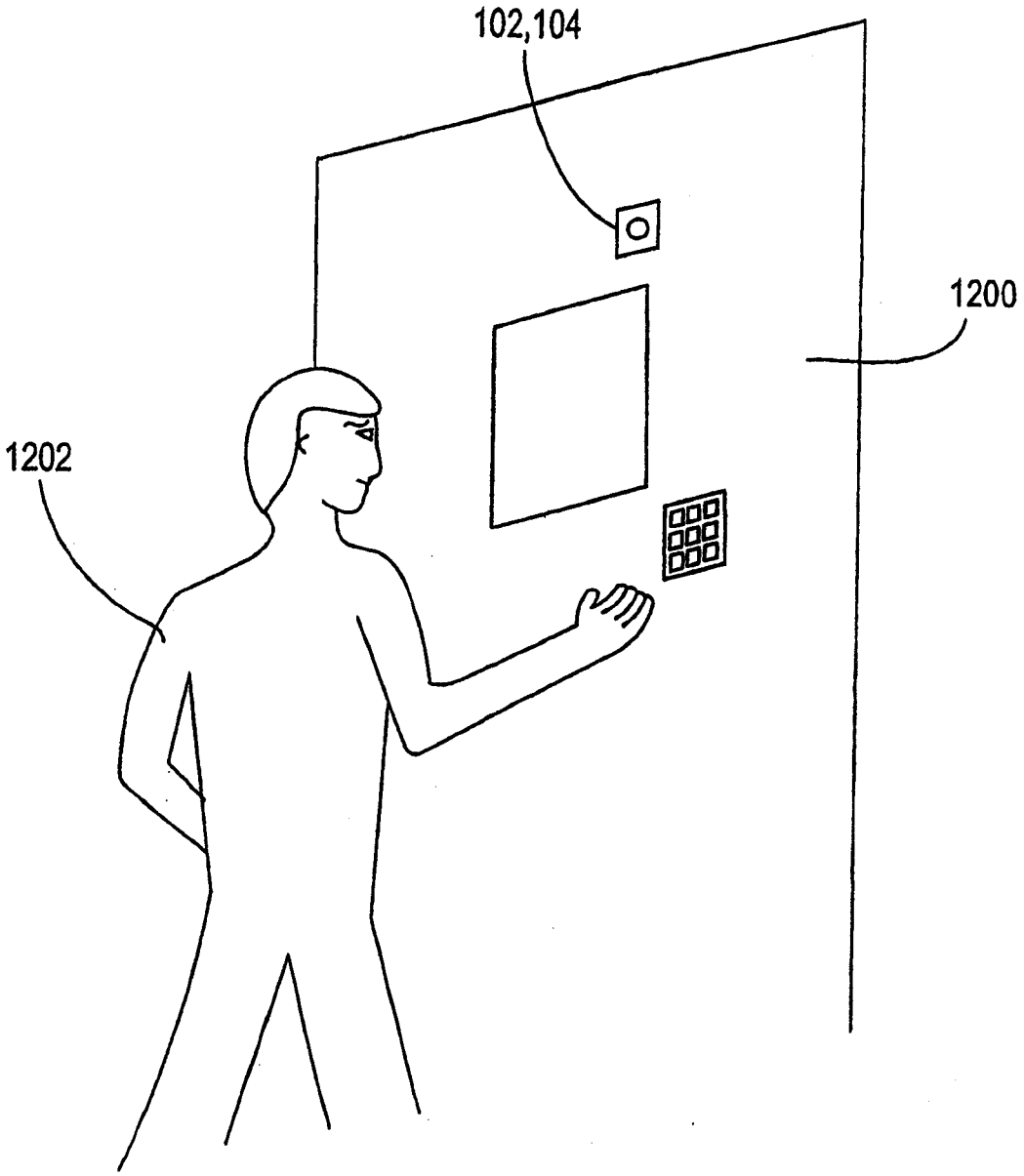


图 12

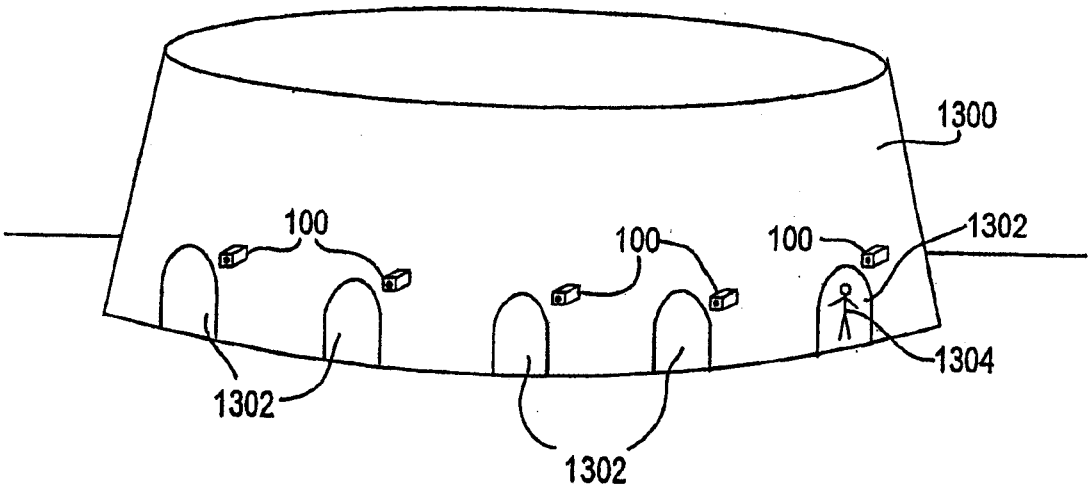


图 13

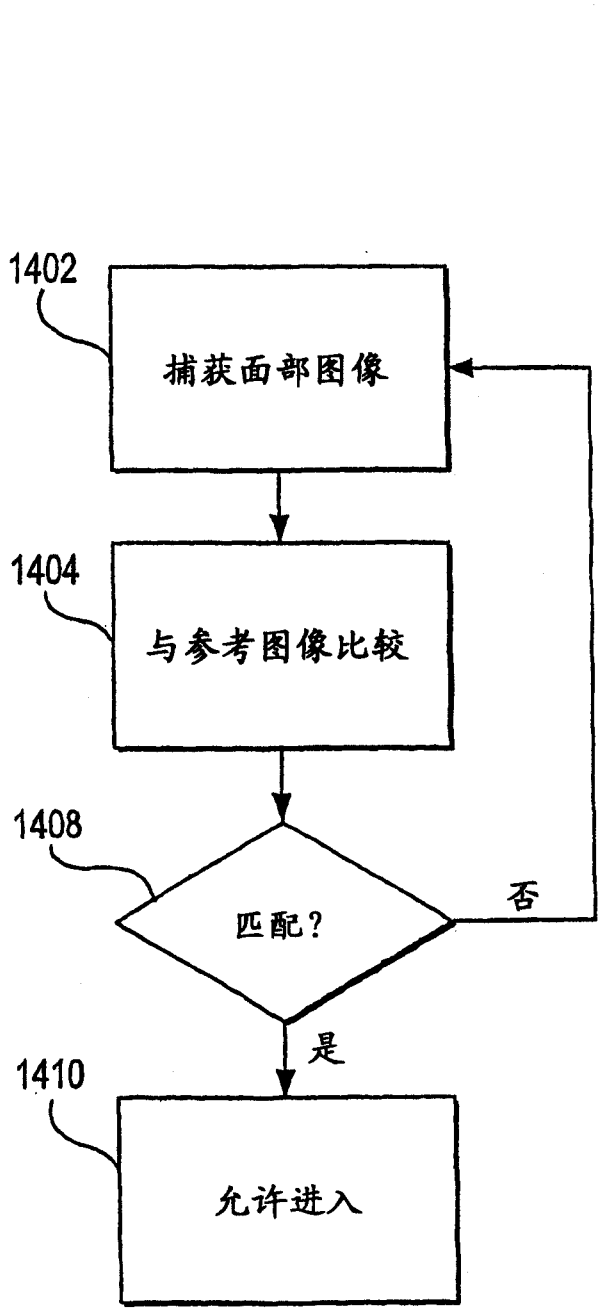


图 14

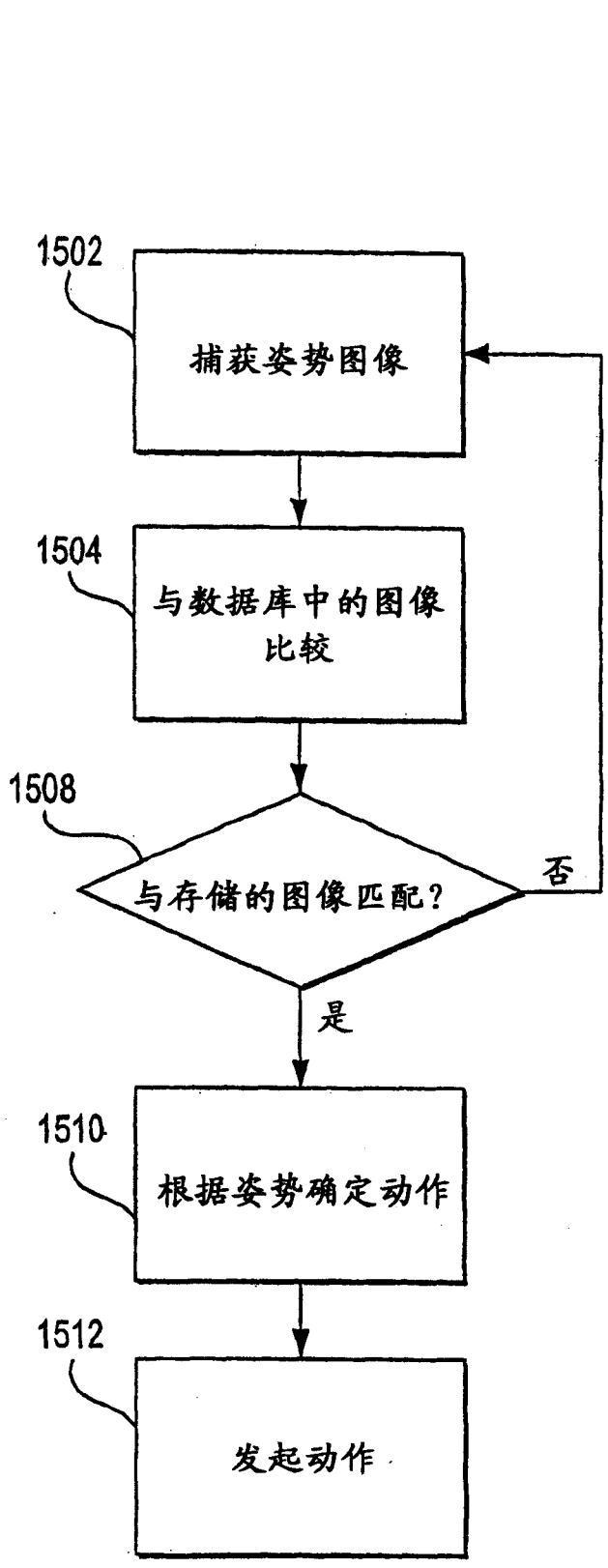

1500
1502
捕获姿势图像
1504
与数据库中的图像比较
1508
与存储的图像匹配?
否
是
1510
根据姿势确定动作
1512
发起动作

图 15

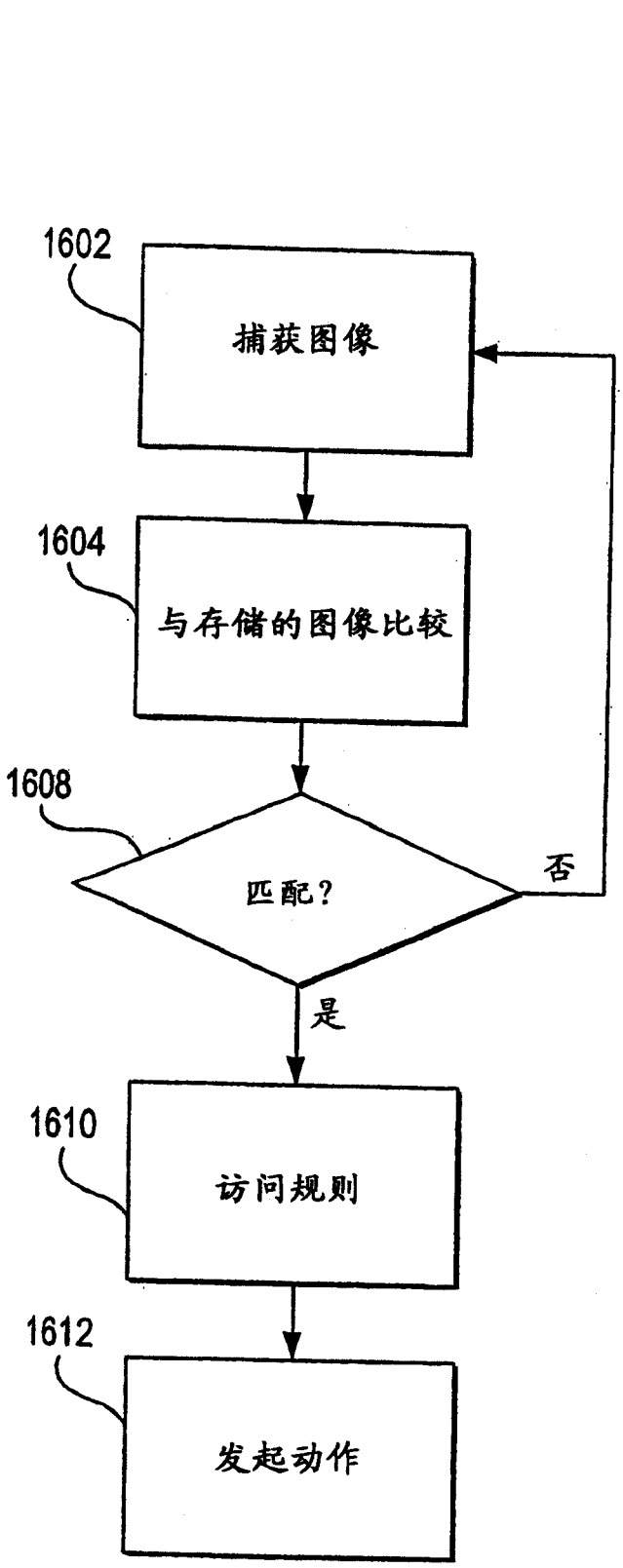


图 16

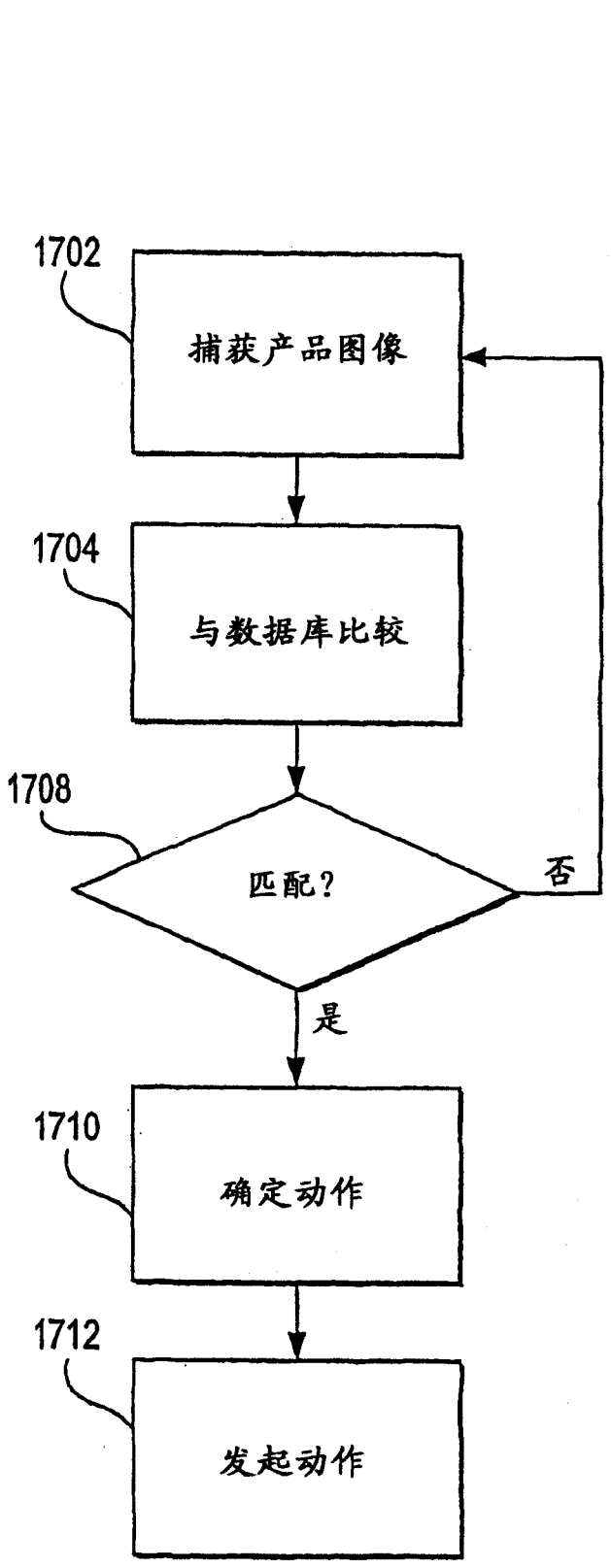


图 17

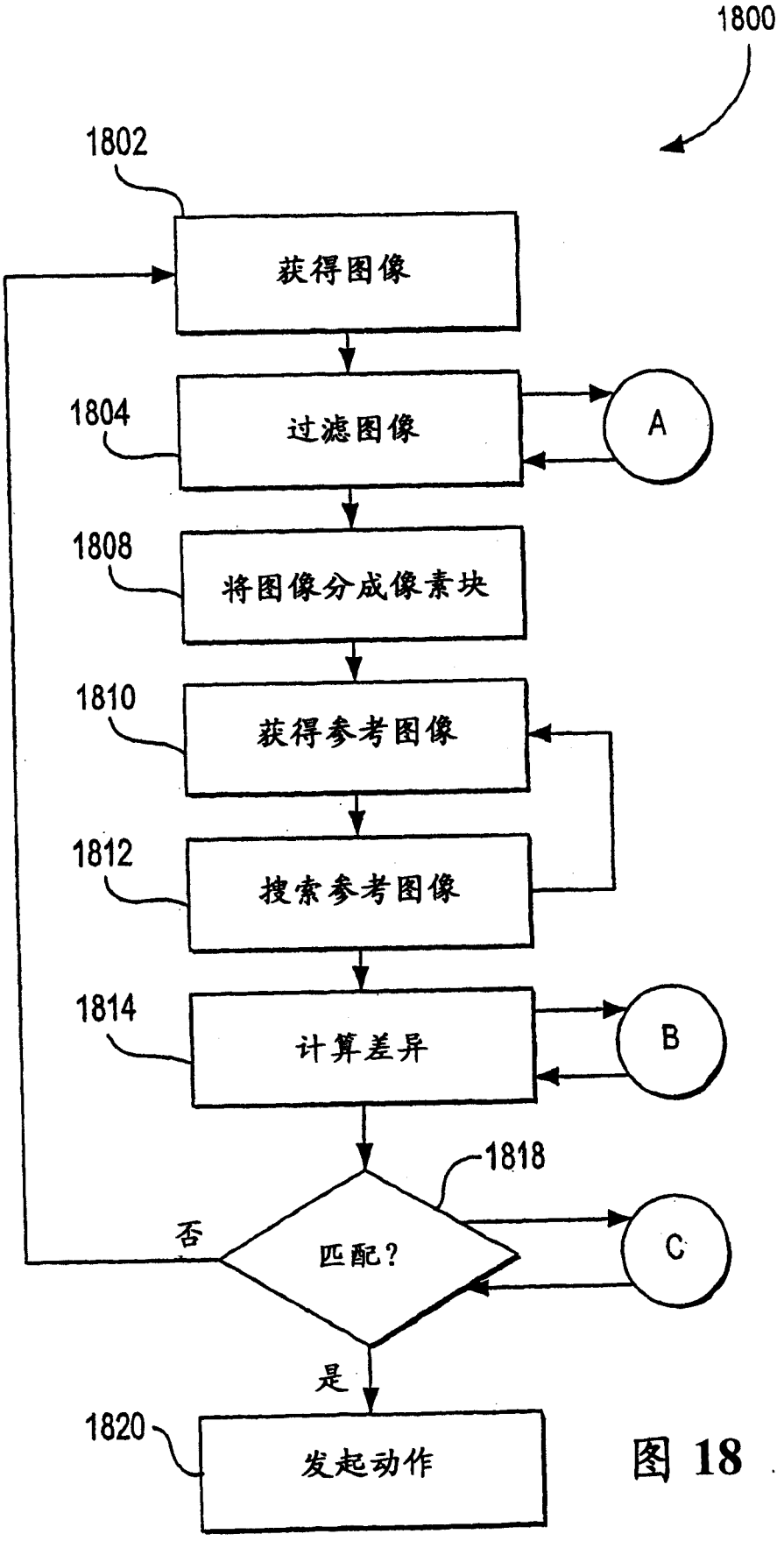

1800
1802
获得图像
1804
过滤图像
A
1808
将图像分成像素块
1810
获得参考图像
1812
搜索参考图像
1814
计算差异
B
1818
否
匹配?
C
是
1820
发起动作

图 18

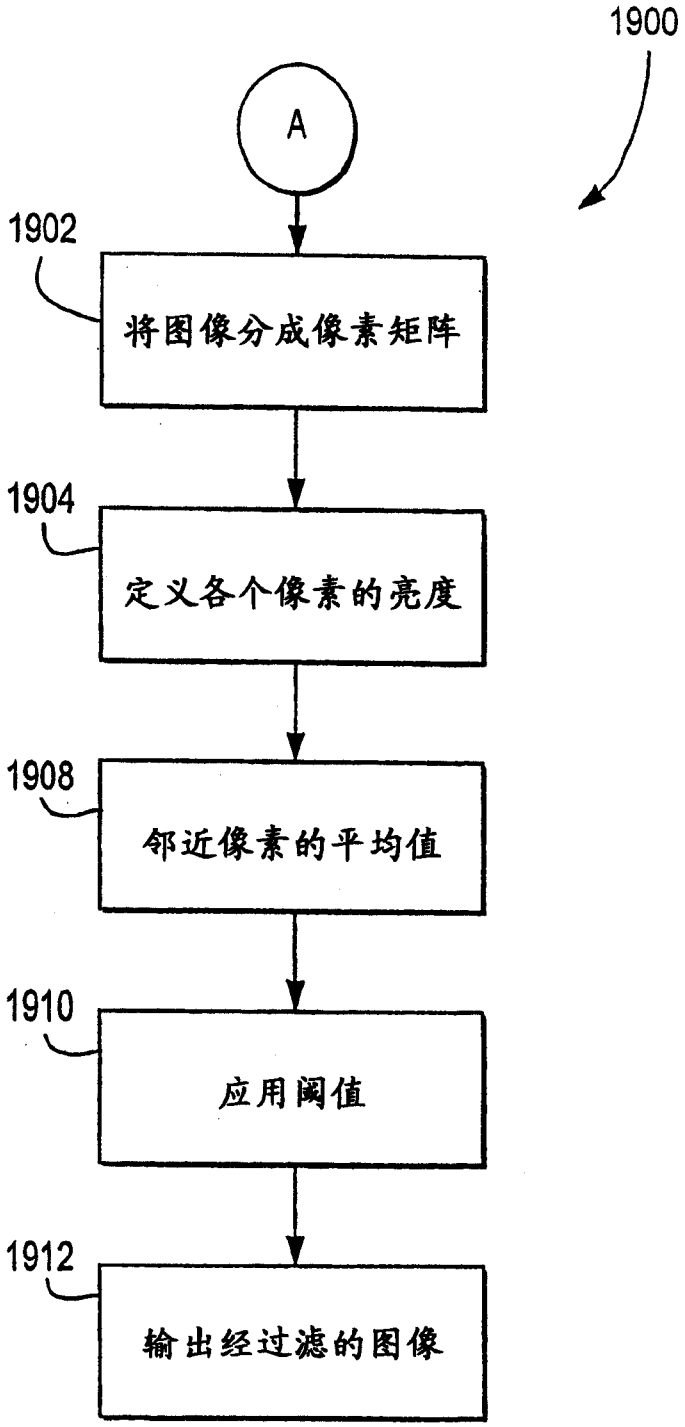

1900
A
1902
将图像分成像素矩阵
1904
定义各个像素的亮度
1908
邻近像素的平均值
1910
应用阈值
1912
输出经过滤的图像

图 19

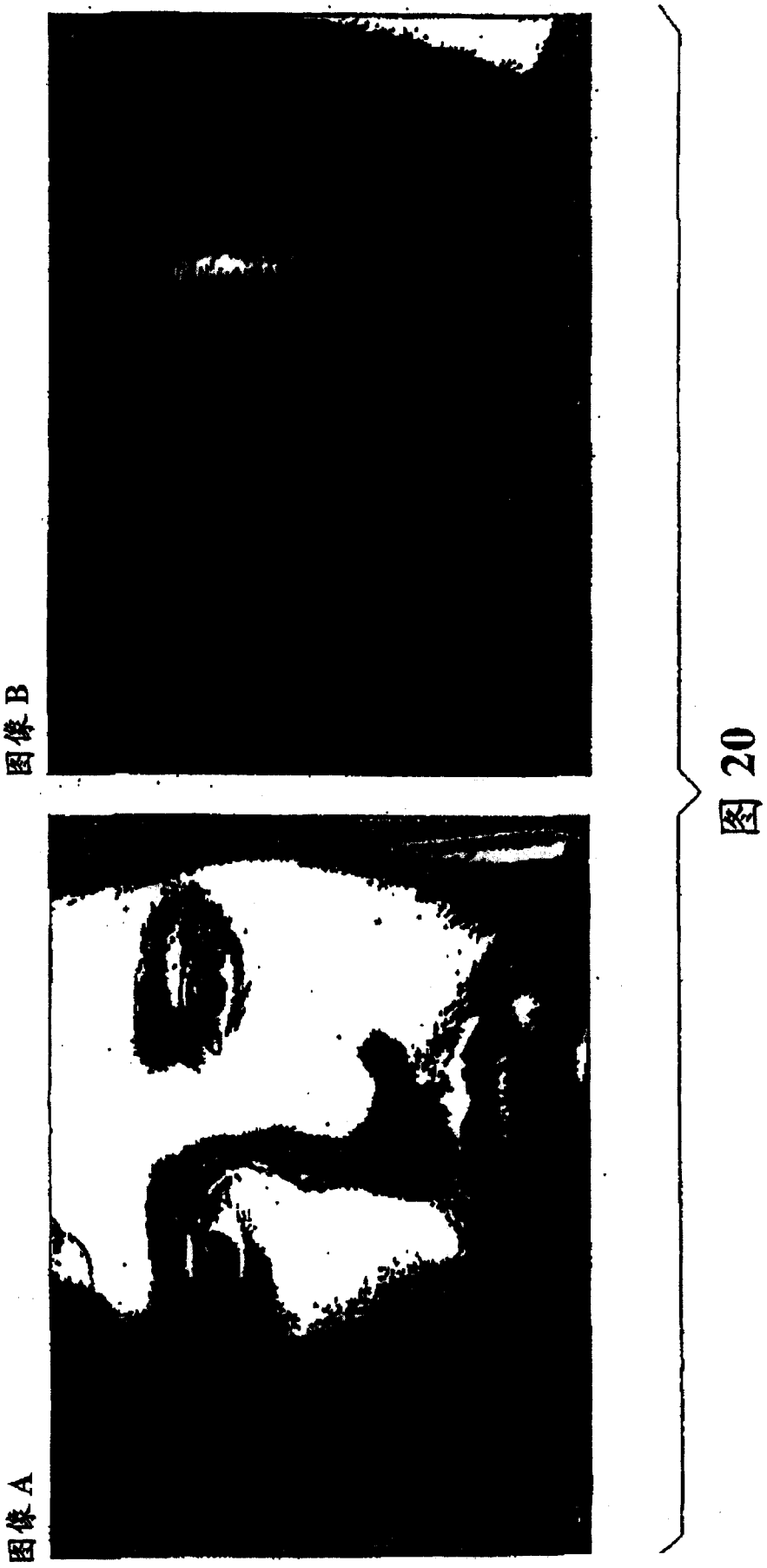


图 20

图 21

面像 A



面像 B



将两幅图像重构。方块表示已接受了相似性的像素块。





在此情况下，成功的相似性的数量是 20 。就实验过程中所采用的搜索参数和允差面积参数来说，这两面部不同的概率小于 $1/10^{10}$ 。

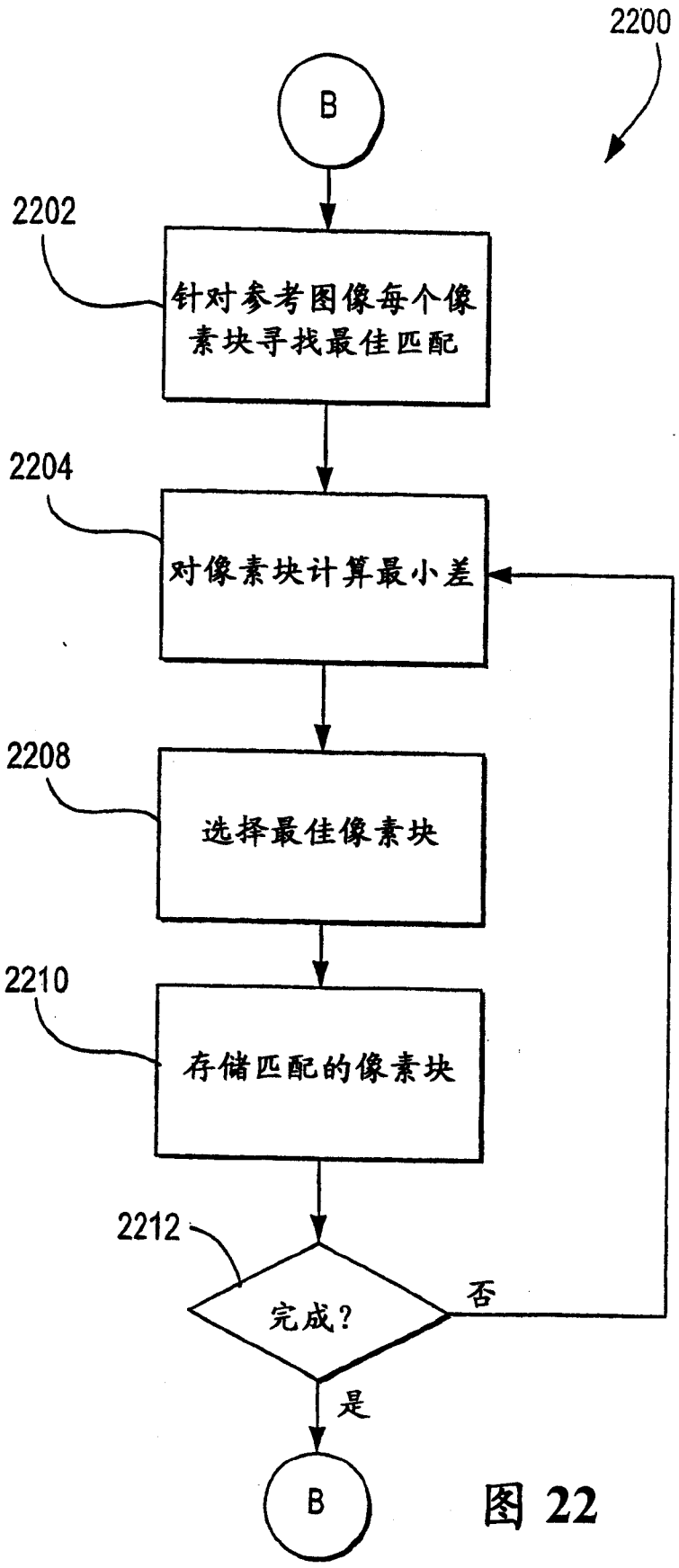

2200
B
2202
针对参考图像每个像素块寻找最佳匹配
2204
对像素块计算最小差
2208
选择最佳像素块
2210
存储匹配的像素块
2212
完成?
否
是
B

图 22

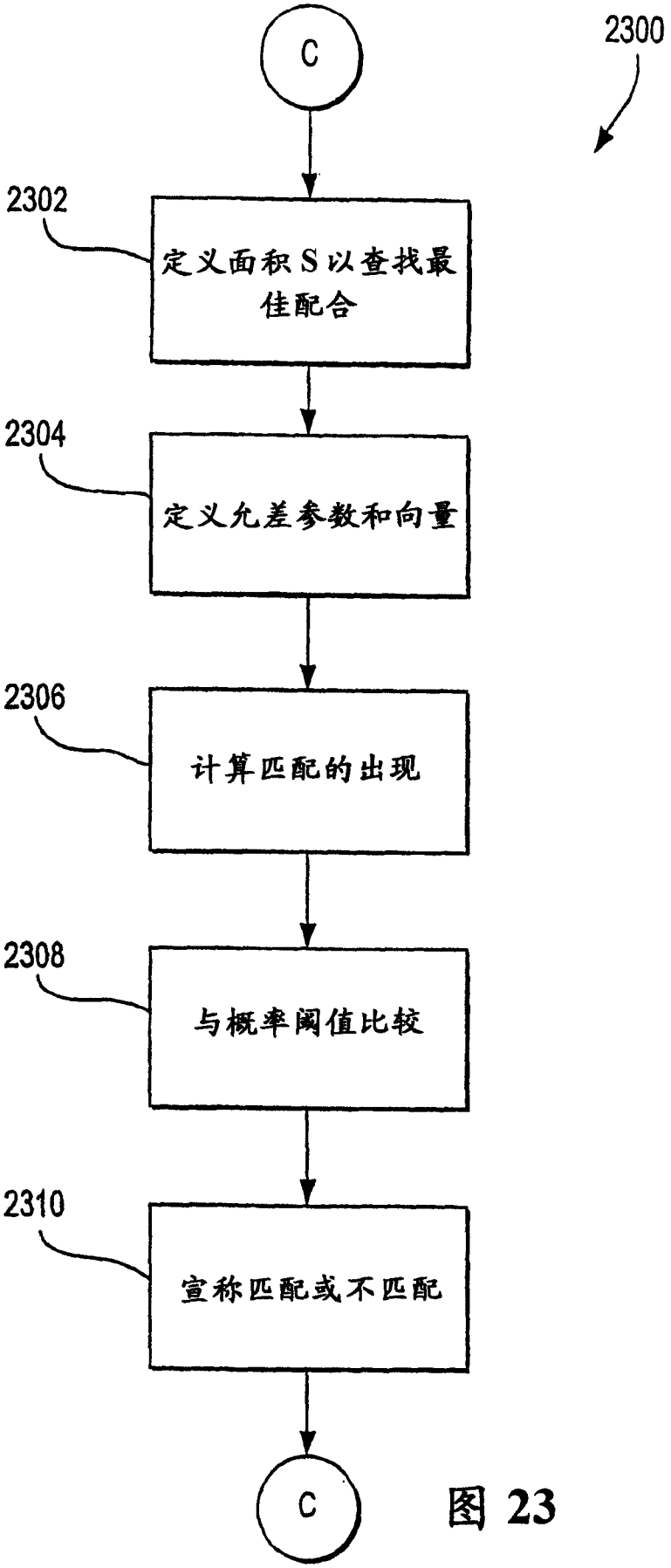

C
2300
2302
定义面积S以查找最佳配合
2304
定义允差参数和向量
2306
计算匹配的出现
2308
与概率阈值比较
2310
宣称匹配或不匹配
C

图 23