



[12] 发明专利申请公开说明书

[21] 申请号 03814775.0

[43] 公开日 2005 年 8 月 31 日

[11] 公开号 CN 1662922A

[22] 申请日 2003.6.13 [21] 申请号 03814775.0
[30] 优先权
[32] 2002. 6. 27 [33] US [31] 10/183,759
[86] 国际申请 PCT/IB2003/002951 2003.6.13
[87] 国际公布 WO2004/003802 英 2004.1.8
[85] 进入国家阶段日期 2004.12.24
[71] 申请人 皇家飞利浦电子股份有限公司
地址 荷兰艾恩德霍芬
[72] 发明人 S·古塔 A·科尔梅纳雷兹
M·特拉科维克

[74] 专利代理机构 中国专利代理(香港)有限公司
代理人 刘红 张志醒

权利要求书 3 页 说明书 7 页 附图 3 页

[54] 发明名称 通过视觉和语音识别测量内容分级
[57] 摘要

提供用于测量客户对至少服务、产品和内容之一的满意度的方法。该方法包括：获得客户的至少图像和语音数据之一；分析获得的至少图像和语音数据之一，用于至少以下之一：(a)检测客户的凝视；(b)检测客户的面部表情；(c)检测客户的情绪；(d)检测客户的语音；以及(e)检测客户与至少服务、产品和内容之一的互相作用；和基于至少(a)-(e)之一来确定客户满意度。

ISSN 1008-4274

1. 一种用于测量客户对至少服务、产品和服务内容（104）之一的满意度的方法，该方法包括：
- 5 获得客户（108a）的至少图像和语音数据之一；
分析获得的至少图像和语音数据之一，用于至少以下之一：
- (a) 检测客户（108a）的凝视；
 (b) 检测客户（108a）的面部表情；
 (c) 检测客户（108a）的情绪；
10 (d) 检测客户（108a）的语音；以及
 (e) 检测客户（108a）与至少服务、产品和服务内容（104）之一的互相作用；和
 基于至少（a）-（e）之一来确定客户满意度。
2. 权利要求1的方法，进一步包括根据至少图像和语音数据之一来确定客
15 户（108a）的至少性别、种族和年龄之一。
3. 权利要求1的方法，其中获得包括在图像数据中识别客户（108a）。
4. 权利要求3的方法，其中识别包括在图像数据中检测面部。
5. 权利要求3的方法，其中识别包括把图像数据中的目标按人和非人来分类。
- 20 6. 权利要求1的方法，其中对客户（108a）的凝视的检测包括确定至少以下之一：检测到的凝视的方向是否对着至少服务、产品和服务内容（104）之一，以及对着至少服务、产品和服务内容（104）之一凝视的持续时间。
7. 权利要求1的方法，其中对客户（108a）的面部表情的检测包括确定检测到的面部表情是满意还是不满意。
- 25 8. 权利要求6的方法，进一步包括：当面部表情被检测到时，检测客户（108a）的凝视是否对着至少服务、产品和服务内容（104）之一，并且其中客户满意度的确定至少部分地基于此。
9. 权利要求1的方法，其中对客户（108a）的情绪的检测至少部分地基于对客户（108a）的至少语音和面部表情之一的检测。
- 30 10. 权利要求1的方法，其中对客户（108a）的情绪的检测包括对客户（108a）

的情绪强度的检测。

11. 权利要求 10 的方法，其中对情绪强度的检测至少部分地基于对客户（108a）的至少语音和面部表情之一的检测。

12. 权利要求 1 的方法，其中对客户（108a）的语音的检测包括检测识别
5 语音的特定短语。

13. 权利要求 1 的方法，其中对客户（108a）的语音的检测包括检测识别语音中的情绪。

14. 权利要求 1 的方法，其中检测客户（108a）与至少服务、产品和服务和内容（104）之一的互相作用包括检测与至少产品、服务和内容（104）之一的物理
10 互相作用。

15. 一种计算机程序产品，在计算机可读媒体中进行实施，用于测量客户对至少服务、产品和服务和内容（104）之一的满意度，该计算机程序产品包括：

用于获得客户（108a）的至少图像和语音数据之一的计算机可读程序代码装置；

15 用于分析获得的至少图像和语音数据之一的计算机可读程序代码装置，以便进行至少以下之一：

（a）检测客户（108a）的凝视；

（b）检测客户（108a）的面部表情；

（c）检测客户（108a）的情绪；

20 （d）检测客户（108a）的语音；以及

（e）检测客户（108a）与至少服务、产品和服务和内容（104）之一的互相作用；

和

用于基于至少（a）-（e）之一来确定客户满意度的计算机可读程序代码装置。

25 16. 权利要求 15 的计算机程序产品，进一步包括用于根据至少图像和语音数据之一来确定客户（108a）的至少性别、种族和年龄之一的计算机可读程序代码装置。

17. 一种机器可读的程序存储设备，有形地实施可由机器执行的指令程序，以执行用于测量客户对至少服务、产品和服务和内容（104）之一的满意度的方法步骤，该方法包括：
30

获得客户（108a）的至少图像和语音数据之一；

分析获得的至少图像和语音数据之一，以便进行至少以下之一：

- 5 (a) 检测客户（108a）的凝视；
- (b) 检测客户（108a）的面部表情；
- (c) 检测客户（108a）的情绪；
- (d) 检测客户（108a）的语音；以及
- (e) 检测客户（108a）与至少服务、产品和内容（104）之一的互相作用；

和

基于至少（a）-（e）之一来确定客户满意度。

- 10 18. 权利要求 17 的程序存储设备，其中该方法进一步包括根据至少图像和语音数据之一来确定客户（108a）的至少性别、种族和年龄之一。

19. 一种用于测量客户对至少服务、产品和内容（104）之一的满意度的设备（100），该设备包括：

- 15 至少照相机（102）和话筒（106）之一，用于获得客户（108a）的至少图像和语音数据之一；和

处理器（114），具有用于分析获得的至少图像和语音数据之一的装置（110，112），用于进行至少以下之一：

- 20 (a) 检测客户（108a）的凝视；
- (b) 检测客户（108a）的面部表情；
- (c) 检测客户（108a）的情绪；
- (d) 检测客户（108a）的语音；以及
- (e) 检测客户（108a）与至少服务、产品和内容（104a）之一的互相作用；

- 25 其中该处理器（114）进一步具有用于基于至少（a）-（e）之一来确定客户满意度的装置。

20. 权利要求 19 的设备，其中该处理器（114）进一步具有用于根据至少图像和语音数据之一来确定客户（108a）的至少性别、种族和年龄之一的装置。

通过视觉和语音识别测量内容分级

5 本发明一般涉及视觉和语音识别，并且尤其涉及用于通过视觉和/或语音识别来测量客户满意度（satisfaction）的方法和设备。

在现有技术中，具有几种已知方式来由客户评价显示的产品、服务或内容（在此共同称为“产品”）的兴趣。然而，所有这些已知方式都是人工完成的。例如，调查卡在产品附近可以得到，以便过路人获取并填写。或者，商店营业
10 员或销售代表可以通过向客户询问一系列与产品有关的问题来征求客户对该产品的兴趣。然而，在任何一种方式中，人们必须愿意参与提问。如果愿意的话，人工提问花费时间，时常比人们愿意花费的时间多得多。此外，人工提问取决于参与人群的诚实性。对于内容，比如电视节目，一项服务 Nielson 自动测量目前正在观看什么内容以及谁在观看。然而，却不自动测量各个人是喜欢还是
15 不喜欢此内容。

此外，显示产品的制造商和销售商常常需要不愿泄露给参与者的信息，比如像性别和种族这样的特征。这类信息对制造商和销售商在市场上出售他们的产品是非常有用的。然而，因为制造商感觉到参与者不想提供这样的信息或者被这样的提问所冒犯，所以制造商和销售商不在他们的产品调查表上提问这样
20 的问题。

因此，本发明的一个目的是提供用于自动测量客户对产品、服务或内容的满意度的方法和设备。因此，提供用于测量客户对至少服务、产品和内容之一的满意度的方法。该方法包括：获得客户的至少图像和语音数据之一；分析获得的至少图像和语音数据之一，用于至少以下之一：（a）检测客户的凝视；
25 （b）检测客户的面部表情；（c）检测客户的情绪；（d）检测客户的语音；以及（e）检测客户与至少服务、产品和内容之一的互相作用；并且基于至少（a）-（e）之一来确定客户满意度。

最好，该方法进一步包括根据至少图像和语音数据之一来确定客户的至少性别、种族和年龄之一。获得最好包括在图像数据中识别客户。识别最好包括
30 在图像数据中检测面部。或者，识别包括把图像数据中的目标按人和非人来分

类。对客户凝视的检测最好包括确定至少检测到的凝视的方向是否对着至少服务、产品和内容之一以及对着至少服务、产品和内容之一的凝视的持续时间。

最好，对客户面部表情的检测包括确定检测到的面部表情是满意还是不满意。该方法最好进一步包括每当面部表情被检测到时检测客户的凝视是否对着至少服务、产品和内容之一，以及其中客户满意度的确定至少部分地基于此。

最好，对客户情绪的检测至少部分地基于对客户的至少语音和面部表情之一的检测。对客户情绪的检测最好包括对客户的情绪强度的检测。

最好，对情绪强度的检测至少部分地基于对客户的至少语音和面部表情之一的检测。对客户语音的检测最好包括在识别出的语音中检测特定短语。

10 最好，对客户语音的检测包括在识别出的语音中检测情绪。

检测客户与至少服务、产品和内容之一的互相作用最好包括检测与至少产品、服务和内容之一的物理互相作用。

还提供用于测量客户对至少服务、产品和内容之一的满意度的设备。该设备包括：至少照相机和话筒之一，用于获得客户的至少图像和语音数据之一；
15 和处理器，具有用于分析获得的至少图像和语音数据之一的装置，用于至少以下之一：（a）检测客户的凝视；（b）检测客户的面部表情；（c）检测客户的情绪；（d）检测客户的语音；以及（e）检测客户与至少服务、产品和内容之一的互相作用；其中该处理器进一步具有用于基于至少（a）-（e）之一来确定客户满意度的装置。

20 最好，该处理器进一步具有用于根据至少图像和语音数据之一来确定客户的至少性别、种族和年龄之一的装置。还提供了用于执行本发明方法的计算机程序产品和用于在其中存储计算机程序产品的程序存储设备。

参考下面的描述、所附权利要求和附图，本发明的设备和方法的这些和其他特点、方面和优点将变得更好理解，其中：

25 图 1 图示用于执行本发明方法的设备的优选实施方案的示意图。

图 2a 和 2b 图示表示本发明方法的优选实施方案的流程图。

现在参考图 1，示出了用于测量客户对至少服务、产品和内容之一的满意度的设备，该设备一般利用参考数字 100 来指示。设备 100 包括至少一个并且最好包括几个照相机 102，具有足以在显示的产品、服务或内容 104 的预定区域
30 域内捕捉图像数据的视野。术语照相机在其通用意义上被用于指所有图像捕捉

设备。照相机 102 最好是数字摄像机, 然而, 它们也可以是模拟摄像机、数字静止图像照相机等。如果使用模拟照相机, 则其输出必须被适当地转换成数字格式。该照相机 102 可以是固定的或者具有随动拍摄、俯摄和推拉摄影性能。该设备还包括至少一个话筒 106, 用于捕捉来自预定区域的语音数据。话筒 106 最好是数字话筒, 然而, 还能够利用其他类型的话筒, 如果其输出信号被适当地转换成数字格式。术语话筒在其通用意义上被用于指所有声音捕捉设备。

照相机 102 和话筒 106 在获得预定区域内的客户 108a、108b 或其他目标 109 的图像和语音数据上是有用的。虽然, 或话筒 106 或至少一个照相机 102 对于实践本发明的方法来说是必需的, 但最好是两者都使用。如在本文中所使用的, 术语“客户”指在照相机 102 和话筒 106 的视野/声音区内图像和/或语音数据中检测到的任何人。客户可能对显示的产品、服务和/或内容感兴趣也可能不感兴趣, 他或她在预定区域内的出现就是足以被分类为“客户”的原因。

捕捉到的图像和语音数据分别利用相应的图像和语音识别装置 110、112 以下面将讨论的方式来分析。设备 100 还包括处理器 114, 比如个人计算机。尽管在图 1 中表示为独立的模块, 但图像和语音识别装置 110、112 最好在处理器 114 中实施, 以执行一套指令, 这套指令分析来自照相机 102 和话筒 106 的输入图像和语音数据。

最好, 处理器 114 进一步具有用于从捕捉到的图像和/或语音数据中确定客户 108a、108b 的至少性别、种族和年龄之一的装置。设备 100 还包括用于输出由处理器 114 分析的结果的输出装置 116。输出装置 116 可以是打印机、监视器或者用在其他方法或设备中的电子信号。

现在将参考图 2a 和 2b 说明本发明的一种方法的优选实施方案。图 2a 和 2b 图示说明了一种方法的优选实施方案的流程图, 该方法最好由设备 100 来执行, 该方法一般利用参考数字 200 来指示。方法 200 测量客户对至少服务、产品和内容(在此一起称为“产品”)之一的满意度。产品可以被显示在公共区域中, 比如商业区, 其中产品(例如, 消费产品)显示在预定区域内, 或者显示在专用区域中, 其中产品(例如, 像电视节目这样的内容)正在预定区域内被观看。

在步骤 202 上, 利用照相机 102 和/或话筒 106 获得预定区域的图像和语音数据中的至少一个并且最好是两者。在获得图像和/或语音数据之后, 在步骤

204 上, 在图像和/或语音数据中识别客户 108a、108b。虽然可以利用图像和语音数据之一或两者来识别预定区域内的客户, 但最好使用本领域中任何已知的方法来这样利用图像数据在图像数据中识别别人。

一种这样的方法是其中在图像数据中检测面部, 并且每个面部与一个人相关。一旦发现一个面部, 那么就可以安全地假定一个人存在。通过检测面部在图像数据中识别人的例子公开在 Gutta 等人的 Mixture of Experts for Classification of Gender, ethnic Origin, and Pose of Human Faces, IEEE Transactions on Neural Networks, Vol. 11, No. 4, July 200 (用于对性别、种族起源和人类面部姿态进行分类的专家系统的混合, 有关神经网络的 IEEE 学报, 第 11 卷, 第 4 号, 7 月 200)。

另一种方法是把图像数据中的目标分为人和非人的。例如, 图 1 中的人 108a、108b 将被分类为客户, 而狗 109 会被分类为非人并且为了分析起见而被删除。这样一个系统的例子公开在 2001 年 2 月 27 日提交的 Gutta 等人的共同未决美国专利申请系列号为 09/794,443 标题为 Classification of Objects through Model Ensembles (通过模型总体来分类目标)。

一旦确定一个人存在, 则可以确定其他特点, 类似于性别、种族起源、面部姿态、面部表情等等。如下面所论述的, 这些特点可以被用于确定客户对所显示产品的兴趣的度量。用于估计一个人的性别和种族起源的方法是本领域中众所周知的, 比如公开在 Gutta 等人的 Mixture of Experts for Classification of Gender, ethnic Origin, and Pose of Human Faces, IEEE Transactions on Neural Networks, Vol. 11, No.4, July 200 (用于对性别、种族起源和人面部姿态进行分类的专家系统的混合, 有关神经网络的 IEEE 学报, 第 11 卷, 第 4 号, 7 月 200)。

通过分析图像和/或语音数据可以确定的一些特点的例子是: 检测客户 108a、108b 的凝视; 检测客户 108a、108b 的面部表情; 检测客户 108a、108b 的情绪; 检测客户 108a、108b 的语音; 以及检测客户 108a、108b 与产品的互相作用, 可以利用其中的一个或多个来测量客户对一个产品的兴趣/满意度。

对于客户 108a、108b 的凝视的检测, 最好在步骤 206 上执行此检测。在步骤 208 上, 最好确定检测到的凝视是否对着产品 104。例如, 图 1 中的客户 108a 将被分类为具有对着产品 104 的凝视, 而客户 108b 将被分类为具有从产品 104 离开的凝视。如果检测到的客户 208b 被发现具有从产品 104 离开的凝

视, 方法 200 沿着路径 208-NO (否) 继续, 并且除了他或她对产品 104 明显不感兴趣, 在分析中不使用客户 108b, 并且该方法循环回到步骤 204, 其中继续在图像数据中识别客户。如果发现客户 108a 具有对着产品 104 的凝视, 则该方法沿着路径 208-YES (是) 继续, 其中检测客户 108a 的其他特点。

- 5 连同凝视的方向一起, 也可以从图像数据中检测出凝视的持续时间, 尤其是对着该产品的凝视的持续时间。可以假定, 对着该产品凝视的持续时间表示对该产品的兴趣。用于在图像数据中检测凝视的方法是本领域中众所周知的, 比如公开在 Rickert 等人的 Gaze Estimation using Morphable Models, Proceedings of the Third International Conference on Automatic Face and Gesture Recognition, Nara, Japan(使用可成形模型进行凝视估算, 在日本奈良举行的有关自动面部和
- 10 手势识别的第三次国际会议的会刊), 1998 年 4 月 14-16 日。

至于客户的面部表情的检测, 最好在步骤 210 上只对于被发现正对着产品 104 凝视的那些客户 108a 来执行此检测。最好, 对客户 108a 的面部表情的检测包括确定检测的面部表情是满意还是不满意之一。例如, 检测到微笑或兴奋

15 神态就表示满意, 而检测到皱眉或茫然不知所措神态就表示不满意。用于检测面部表情的方法是本领域中众所周知的, 比如公开在 Colmenarez 等人的 Modeling the Dynamics of Facial Expressions, CUES Workshop held in conjunction with the International Conference on Computer Vision and Pattern Recognition, Hawaii, USA, December 10-15, 2001 (给面部表情的动态建模, 2001 年 12 月 10-

20 15 日在美国夏威夷与计算机视觉和模式识别有关的国际会议一起举行的 CUES 专题讨论会中公开的)。

至于语音的检测, 最好在步骤 212 上执行此检测, 并且这不仅对于在预定区域内识别客户 108a、108b, 而且对于确定他们对该产品的满意度的度量都是有用的。例如, 对客户 108a、108b 的语音的检测可以在识别语音中检测特定

25 短语。例如, 术语“好极了”或“酷”的识别将表示满意的度量, 而术语“真臭”或“糟透了”将表示不满意的度量。

在步骤 214 上, 可以对检测到的客户 108a、108b 的情绪进行检测。因为客户 108a 正在凝视该产品, 所以将只检测到他或她的情绪。客户 108a 的情绪的检测最好(至少部分地)基于客户 108a 的语音和/或面部表情的检测。此外,

30 还可以对检测到的情绪的强度进行检测。例如, 某个面部表情比如兴奋的神色

具有比微笑更大的情绪强度。类似地，还可以在检测到的客户 108a 的语音中检测情绪的强度，比如此客户改变他的语音模式（例如，说得更快或更大声）或者使用惊叹语。在面部表情和语音中情绪的识别是本领域中众所周知的，比如公开在下述文献中的内容：Colmenarez 等人的 Modeling the Dynamics of Facial Expressions, CUES Workshop held in conjunction with the International Conference on Computer Vision and Pattern Recognition, Hawaii, USA, December 10-15, 2001（2001 年 12 月 10-15 日在美国夏威夷与计算机视觉和模式识别有关的国际会议一起举行的 CUES 专题讨论会）；以及 Frank Dellaert 等人的 Recognizing Emotion in Speech, in Proc. of Int'l Conf. On Speech and Language Processing(1996)(识别语音中的情绪，有关语音和语言处理的国际会议（1996）的会刊）；以及 Polzin 等人的 Detecting Emotions in Speech, Proceedings of the Cooperative Multimodal Communication Conference, 1998（检测语音中的情绪，1998 年协作多形式通信会议的会刊）。

在步骤 216 上，确定是否有客户 108a 与产品 104 的互相作用，比如与该产品的物理互相作用。例如，对于被显示的产品（例如，汽车），对客户 108a 触摸该产品以及有可能玩弄该产品的某些开关或其他部分的确定可以表示对该产品的满意度的度量，尤其在外加检测到赞成的情绪、语音和/或面部表情的时候。物理互相作用的确定可以通过分析来自照相机 102 的图像数据和/或来自触觉元件（未示出）的反馈来进行。这种用于确定与产品的物理互相作用的方法是本领域中众所周知的。

如上面所论述的，对客户 108a、108b 的其他特点比如性别、种族起源和年龄的检测还可以最好在步骤 218 上进行。虽然这样的特点在确定对产品满意度的测量上可能不是有用的，但在市场出售方面可以是非常有用的。例如，方法 200 可以确定大多数女人对一个特定产品满意，而大部分男人对此产品不是不满意就是不感兴趣。类似的市场销售策略可以从对满意度和种族起源和/或年龄的分析中获悉。

在步骤 220 上，基于至少上面所论述的特点之一来确定客户满意度，而且最好基于这些特点的组合来确定。用于这样的确定的简单算法是把权重分配给这些特点中的每一个并从其计算一个分数，该分数表示满意/不满意的度量。即，少于预定数的分数将表示对产品 104 不满意，而高于预定数的分数将表示对产

品 104 满意。

另一个例子是给每个特点分配一个表示可能满意的得分，其中所有检测到的特点的累积分数超过预定数将表示对产品 104 满意，而累积分数低于预定数将表示对产品 104 不满意。该算法也可以是复杂的并且用于大量的情景和检测到的特点的组合。例如，如上面所论述的，被检测到正长时间持续凝视着产品 104 并且在他或她的语音和面部表情中检测到高的情绪强度的客户 108a 将表示对该产品极为满意，而带着不满意的面部表情看着产品并且在他或她的语音中具有不满意情绪的客户 108a 将表示对该产品几乎没兴趣或完全没兴趣。类似地，只短时间扫了一眼产品 104 并且在他或她的语音和面部表情中几乎没有或完全没有情绪的客户 108a 可以表示对产品 104 几乎没兴趣或者完全没兴趣。

在步骤 222，该分析的结果被输出用于再观看、满意分析或在其它的方法或设备中使用。

本发明的方法特别适合于利用计算机软件程序来执行，这样的计算机软件程序最好包含与方法的各个步骤相对应的模块。这样的软件当然可以实施在计算机可读媒体比如集成芯片或者外围设备中。

尽管已经示出并描述了被认为是本发明优选实施方案的方案，但是当然将明白，不难进行形式或细节上的各种修改和改变而不脱离本发明的精神。因此，意图本发明不限于所述的和所示的那些确切形式，而应被构造为覆盖可能落在所附权利要求的范畴之内的所有修改。

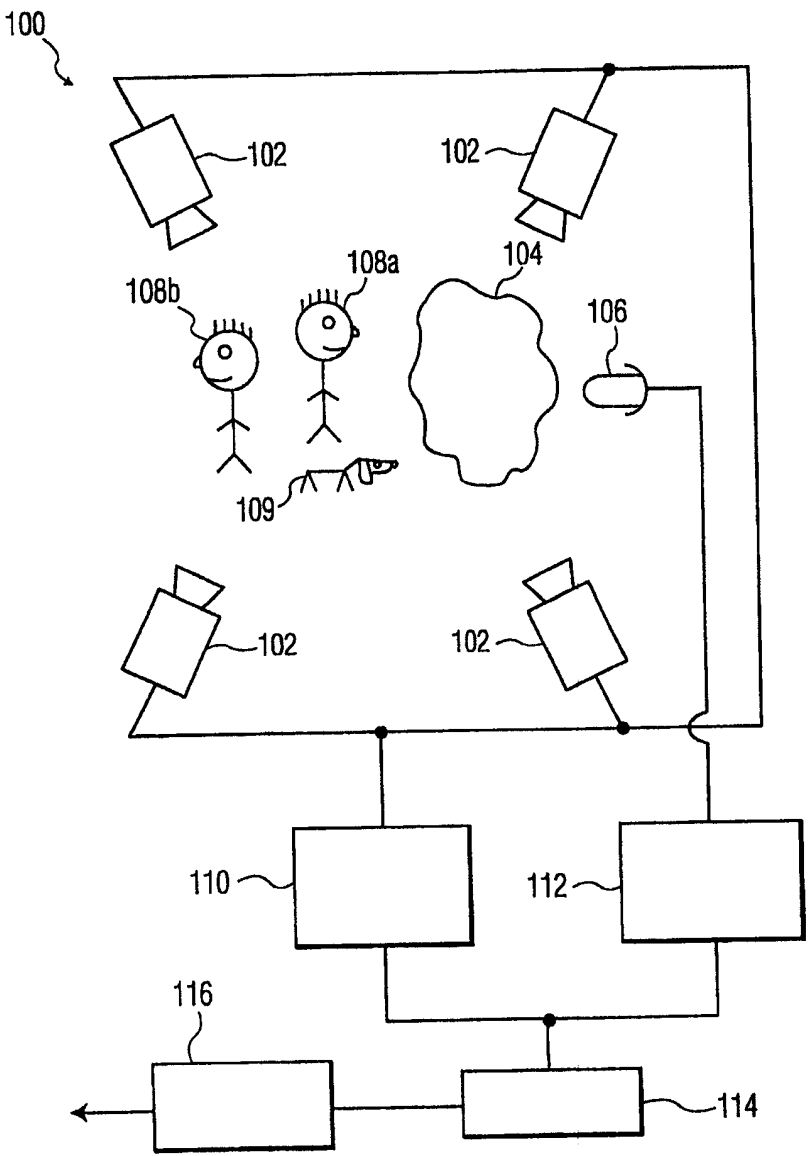


图 1

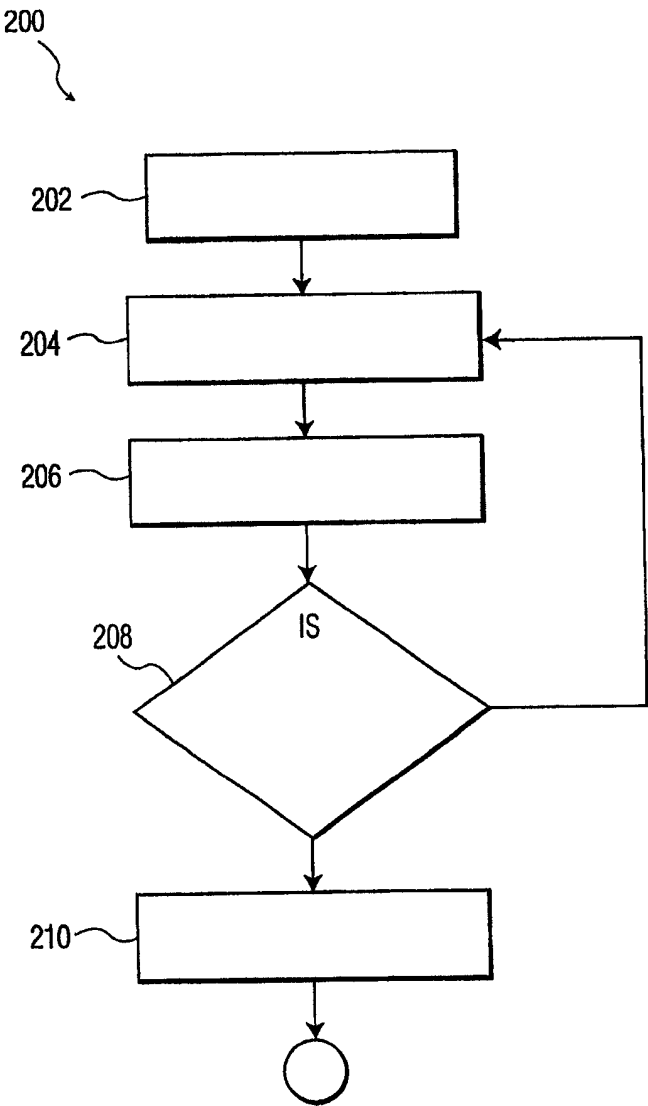


图 2a

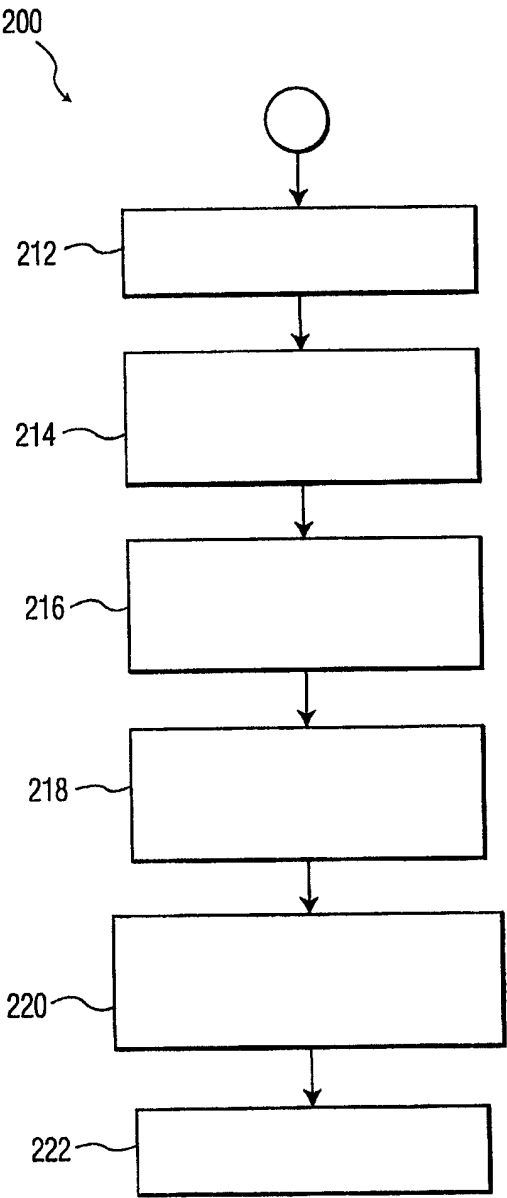


图 2b