



(12) 发明专利申请

(10) 申请公布号 CN 102227753 A

(43) 申请公布日 2011. 10. 26

(21) 申请号 200980147669. 0

(22) 申请日 2009. 10. 01

(30) 优先权数据

61/102, 618 2008. 10. 03 US

(85) PCT申请进入国家阶段日

2011. 05. 27

(86) PCT申请的申请数据

PCT/US2009/059225 2009. 10. 01

(87) PCT申请的公布数据

W02010/039954 EN 2010. 04. 08

(71) 申请人 3M 创新有限公司

地址 美国明尼苏达州

(72) 发明人 布拉因·J·斯坦凯维奇

布赖恩·E·布鲁克斯

布赖恩·L·林兹 纳森·J·安德森

迈克尔·凯利·卡那万

格兰·E·卡斯纳尔

蒂莫西·J·加德纳

戴维·K·麦斯米尔

(74) 专利代理机构 中原信达知识产权代理有限  
责任公司 11219

代理人 周亚荣 安翔

(51) Int. Cl.

G06T 11/60(2006. 01)

G06K 9/46(2006. 01)

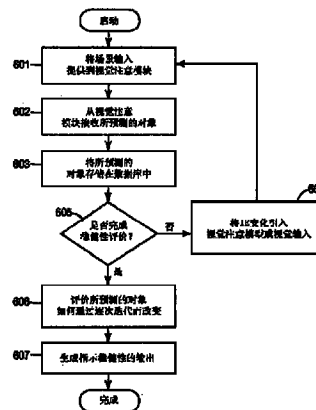
权利要求书 3 页 说明书 25 页 附图 20 页

(54) 发明名称

用于评估稳健性的系统和方法

(57) 摘要

本发明公开了用于评价场景内对象或场景本  
身的稳健性的系统和方法。



1. 一种计算机系统,包括:  
视觉注意模块,其可用来:
  - (a) 接收场景输入;
  - (b) 将视觉注意模型应用到所述场景输入;以及
  - (c) 预测所述场景输入内易于吸引视觉注意的区域;以及稳健性评估模块,其可用来与所述视觉注意模块进行交互以确定经识别的区域中的至少一个的稳健性程度或所述场景的稳健性程度。
2. 根据权利要求 1 所述的系统,其中稳健性是对从视觉注意模型输出的预测对以下任一者或某个组合的灵敏度的度量:
  - (a) 所述场景的所述视觉特性内的变动;或
  - (b) 通过所述视觉注意模块施加的所述视觉注意模型内的变动。
3. 根据权利要求 2 所述的系统,其中所述稳健性评估模块还可用来将变动引入所述场景的视觉特性。
4. 根据权利要求 3 所述的系统,其中所述稳健性评估模块通过修改所述场景输入将变动引入所述场景的视觉特性。
5. 根据权利要求 4 所述的系统,其中修改所述场景输入包括系统地或随机地选择所述图像内的像素来修改,并系统地或随机地扰乱那些所选像素的特性。
6. 根据权利要求 2 所述的系统,其中所述稳健性评估模块还可用来将通过所述视觉注意模块施加的变动引入所述视觉注意模型。
7. 根据权利要求 6 所述的系统,其中通过修改所述视觉注意模型的操作参数将所述变动引入所述视觉注意模型。
8. 根据权利要求 7 所述的系统,其中修改所述操作参数是系统地或随机地进行的。
9. 根据权利要求 7 所述的系统,其中被修改的所述操作参数是控制所述视觉注意模型对以下任一者的灵敏度的那些:颜色、对比度、运动、形状、空间频率、取向。
10. 根据权利要求 2 所述的系统,其中所述稳健性评估模块还可用来同时:
  - (a) 将变动引入所述场景的视觉特性;并
  - (b) 将通过所述视觉注意模块施加的变动引入所述视觉注意模型。
11. 根据权利要求 10 所述的系统,其中将变动引入所述场景的所述视觉特性包括将视觉干扰物引入所述场景。
12. 根据权利要求 10 所述的系统,还包括:  
数据库模块,其可用来记录所述经识别的区域如何根据引入场景的视觉特性的变动、或引入所述视觉注意模型内的通过所述视觉注意模块施加的变动而变化,所述数据库模块还可用来记录变动的类型。
13. 根据权利要求 12 所述的系统,其中所述稳健性评估模块重复地:
  - (a) 将变动引入所述场景的所述视觉特性,从而生成修改场景,以及;
  - (b) 调用所述视觉注意模块并提供所述修改场景作为输入。
14. 根据权利要求 12 所述的系统,其中所述稳健性评估模块重复地:
  - (a) 将通过所述视觉注意模块施加的变动引入所述视觉注意模型,以及;
  - (b) 调用所述视觉注意模块以使用所述修改的视觉注意模型。

15. 根据权利要求 1 所述的系统,还包括:

用户接口模块,其可用来显示指示所述经识别的区域中的至少一个具有稳健性的程度的信息。

16. 根据权利要求 15 所述的系统,其中所述信息包括所述场景的图形表达,所述场景的所述图形表达覆盖有与所述场景的单独区域相关的热图,所述热图的梯度与所述场景的单独区域的所述稳健性相关联。

17. 根据权利要求 15 所述的系统,其中所述信息包括数字。

18. 根据权利要求 1 所述的系统,其中所述稳健性评估模块另外具有可观察特定区域的持续时间,并另外使用该信息来确定所述区域是否具有稳健性。

19. 根据权利要求 1 所述的系统,其中所述视觉注意模型是经验模型。

20. 根据权利要求 2 所述的系统,其中所述稳健性评估模块还可用来将所述稳健性的程度量化为数字,并将该数字提供到计算机程序。

21. 一种计算机实现的方法,包括:

接收场景输入;

将视觉注意模型应用到所述场景输入以预测所述场景输入内易于吸引视觉注意的区域;

改变所述场景以产生修改的场景;

将所述视觉注意模型应用到所述修改场景输入以预测所述场景输入内易于吸引视觉注意的区域;以及,

基于由将所述视觉注意模型应用到所述场景输入和所述修改的场景而识别的所述区域,确定经识别的区域中的一个或所述场景输入具有稳健性的程度。

22. 根据权利要求 21 所述的计算机实现的方法,其中稳健性是对视觉注意模型的预测对所述场景的所述视觉特性内变动的灵敏度的度量。

23. 根据权利要求 22 所述的计算机实现的方法,其中改变所述场景包括系统地或随机地选择所述图像内的像素来修改,并系统地或随机地扰乱那些所选像素的特性。

24. 根据权利要求 22 所述的计算机实现的方法,其中改变所述场景包括将视觉干扰物引入所述场景。

25. 根据权利要求 22 所述的计算机实现的方法,其中改变所述场景包括修改在所述场景内的包括对象的区域。

26. 根据权利要求 22 所述的计算机实现的方法,其中改变所述场景包括修改在所述场景内的包括对象的区域的尺寸、颜色或形状。

27. 根据权利要求 22 所述的计算机实现的方法,其中改变所述场景包括调节所述场景的亮度或对比度。

28. 根据权利要求 22 所述的计算机实现的方法,还包括:

经由用户接口向用户呈现指示所述经识别的区域中的一个或所述场景输入具有稳健性的程度的信息。

29. 根据权利要求 28 所述的计算机实现的方法,其中所述信息是数字。

30. 根据权利要求 21 所述的计算机实现的方法,还包括:

向计算机程序提供指示所述经识别的区域中的一个或所述场景输入具有稳健性的程

度的数字。

31. 一种计算机实现的方法,包括:

接收场景输入;

将视觉注意模型应用到所述场景输入以预测所述场景输入内易于吸引视觉注意的区域;

改变所述视觉注意模型以产生修改的视觉注意模型;

将所述修改的视觉注意模型应用到所述场景输入以预测所述场景输入内易于吸引视觉注意的区域;以及,

基于由将所述视觉注意模型和所述修改的视觉注意模型应用到所述场景输入而识别的区域,确定所述经识别的区域中的一个或所述场景输入具有稳健性的程度。

32. 根据权利要求 31 所述的计算机实现的方法,其中稳健性是对来自视觉注意模型的预测对所述视觉注意模型内的变动的灵敏度的度量。

33. 根据权利要求 32 所述的计算机实现的方法,其中改变所述视觉注意模型包括修改所述视觉注意模型的操作参数。

34. 根据权利要求 33 所述的计算机实现的方法,其中所述操作参数被随机地或系统地修改。

35. 根据权利要求 34 所述的计算机实现的方法,其中所述操作参数是控制所述视觉注意模型对以下任一者或其组合的灵敏度的那些:颜色、对比度、运动、形状、空间频率和取向。

36. 根据权利要求 31 所述的计算机实现的方法,还包括:

提供指示所确定的稳健性程度的输出。

37. 根据权利要求 36 所述的计算机实现的方法,其中所述输出是数字。

38. 根据权利要求 37 所述的计算机实现的方法,其中所述输出提供给计算机程序。

## 用于评估稳健性的系统和方法

[0001] 相关专利申请的交叉引用

[0002] 本专利申请要求 2008 年 10 月 3 日提交的美国临时专利申请 No. 61/102618 的优先权。

[0003] 相关专利申请

[0004] 本专利申请还涉及 2008 年 10 月 3 日提交的美国临时专利申请 No. 61/102625 和 2008 年 10 月 3 日提交的美国临时专利申请 No. 61/102670。

### 背景技术

[0005] 人类视觉系统是容量有限的系统,因为其在任何特定时间只能处理较少数量的对象。尽管事实上在任何特定时间有许多对象可以观察到,但情况确实如此。人类视觉系统在任何特定时间将只注意或处理来自人类可见对象的阵列中的一个(或极少数)对象。当人观察图像或场景时,其视觉系统将让注意(和心智处理)从一对象转移到另一对象。

[0006] 在人类视觉注意领域已有大量的研究。该项研究已产生许多针对理解人类视觉注意行为的研究成果,以及许多视觉注意的计算模型。这些计算模型(有时称为视觉注意模型、眼球凝视预测模型、注意模型或显著性模型)在给定视觉刺激物(例如,图片或场景)下预测人将让其视觉注意或凝视分配在何处。

[0007] 这些模型提供对场景内将吸引视觉注意的对象或区域的预测。然而,典型的真实世界场景常常是高度动态的。当(例如)人的有利位置改变、场景内的对象改变位置或定向、或者照明改变(投射不同阴影)时,投射到人的图像将随之改变。此外,观察者自身会将不确定性引入预测(观察者会心不在焉或者说是倾向于特定的注意模式)。从场景投射图像的任何变动或观察者的变动、或者甚至场景本身的微小改变都会显著地改变这些模型作出的预测。在将视觉注意模型用于应用装置时会由此而产生问题。

### 发明内容

[0008] 用于评估(以及在一些实施例中量化)场景内对象对场景和观察者变动的稳健性或场景本身对场景和观察者变动的稳健性的系统和方法。特别是,用于将变动引进到场景和/或视觉注意的计算模型、然后评估这些变化对视觉注意将预测分配在何处的影响的方法。

[0009] 在一个实施例中描述了一种具有视觉注意模块和稳健性评估模块的计算机系统,视觉注意模块可用来:(a)接收场景输入、(b)将视觉注意模型应用到场景输入以及(c)预测场景输入内易于吸引视觉注意的区域;稳健性评估模块可用来与视觉注意模块进行交互以确定经识别的区域中的至少一个具有稳健性或场景具有稳健性的程度。

[0010] 在另一个实施例中,描述了一种计算机实现的方法,其中包括:接收场景输入;将视觉注意模型应用到场景输入以预测场景输入内易于吸引视觉注意的区域;改变场景以产生经修改场景;将视觉注意模型应用到经修改场景输入以预测场景输入内易于吸引视觉注意的区域;以及,基于由将视觉注意模型应用到场景输入和被修改的场景而识别的区域,确

定经识别的区域中的一个或场景输入具有稳健性的程度。

[0011] 在另一个实施例中,描述了一种计算机实现的方法,其中包括:接收场景输入;将视觉注意模型应用到场景输入以预测场景输入内易于吸引视觉注意的区域;改变视觉注意模型以产生经修改的视觉注意模型;将经修改的视觉注意模型应用到场景输入以预测场景输入内有助于吸引视觉注意的区域;以及,基于由将视觉注意模型和经修改的视觉注意模型应用到场景输入所辨认的区域,确定经识别的区域中的一个或场景输入具有稳健性的程度。

## 附图说明

[0012] 图 1 是示出代表性视觉注意模块的高级功能的流程图。

[0013] 图 2 是代表性场景的草图。

[0014] 图 3A 是视觉注意模块输出的艺术效果图。

[0015] 图 3B 是视觉注意模块输出的艺术效果图。

[0016] 图 4 是示出两个实例对象的特性的曲线图。

[0017] 图 5 示出广告牌对象可以如何以不同背景场景为背景来设定。

[0018] 图 6 是用于评估场景或场景内对象的稳健性的系统中功能模块的示意图。

[0019] 图 7 是示出实例场景修改的具体说明。

[0020] 图 8 示出视觉注意稳健性评估系统的一个替代实施例,其中视觉注意稳健性评估系统另外包括网络服务器模块。

[0021] 图 9 是说明视觉注意稳健性评估系统评估场景内对象或场景本身的稳健性可采用的流程的一个实施例的高级流程图。

[0022] 图 10 是示出场景优化的高级流程的流程图。

[0023] 图 11 是说明与注意扫视所关注对象的连续位置有关的两种不同回报结构的曲线图。

[0024] 图 12 是场景优化系统中的功能模块的示意图。

[0025] 图 13A 和图 13B 各自为一个场景的艺术效果图。

[0026] 图 14A 和图 14B 各自为一个场景的艺术效果图。

[0027] 图 15 是示出多视角场景分析的高级流程图。

[0028] 图 16 是会议室的示意图。

[0029] 图 17A 和图 17B 示出会议室的不同视角。

[0030] 图 18 是说明对会议室中各种对象进行的三个不同实例分析的图。

[0031] 图 19 是多视角场景分析系统的示意图。

## 具体实施方式

[0032] 视觉注意模型

[0033] 视觉注意模型确定场景内区域在其将吸引视觉注意或吸引眼球运动的可能性方面相异的程度。场景是经受视觉注意模型评估的任何输入形式(例如,图形图像),并可以是(例如)数码照片、虚拟 3D 场景、网页、文件或视频。

[0034] 对象的视觉注意模型倾向是指视觉注意的模型如何表征场景内对象的相对显著

性。例如,一些视觉注意模型将在所预测对象周围叠加迹线。其他视觉注意模型将生成可以在图像上叠加或与图像分开看的热图。还有一些模型可以生成并分配一个值到特定对象和/或区域,该值相对地代表对象的显著性。在迹线的语境中,对象的倾向可以认为是由模型“选定”(加迹线时)或“未选定”。在热图的语境中,对象的倾向是算法已选定对象(或未选定对象)的程度。而在显著性数字的语境中,对象的倾向可以是显著性数字本身。

[0035] 有许多视觉注意模型可预测人类视觉注意会被分配到场景内的何处。通常,这些视觉注意模型采用环境的单幅图像作为输入,并生成对注意将分配在该场景内何处的预测。实验方法使人类受试者接触到场景,并跟踪其眼球移动。然而,该实验方法是资源密集型的,所以已经开发出多个数学模型以通过至少部分地分析场景来预测注意。那就是说实验方法还可以用作本文所述系统和方法的一部分,并如本文所用,实验方法被认为是视觉注意建模的一种类型。

[0036] 图 1 表示这些模型之一的基本方法论,其由 Itti, L. & Koch, C(2000) 在“基于显著性的视觉注意明显和隐蔽转移的搜索机制”(A saliency-based search mechanism for overt and covert shifts of visual attention)(《视觉研究》(Vision Research),第 40 卷,1489-1506 页)中提出。在高标准下,图 1 示出如何通过评估“由下而上”特征(例如,颜色、运动、亮度、边缘等)来进行视觉注意的预测,所述特征用作传达人类视觉的一些方面的视觉表现的构成要素。首先,将数码照片形式的场景提供给 Itti & Koch 模型的计算机实现版本(步骤 10)。接着,特征提取过程分析数码照片的颜色、强度、取向或其他场景线索,例如运动、接点、明暗界线、立体视差和从明暗恢复形状(步骤 11)。特征提取过程产生多幅特征映射(步骤 12),所述特征映射相结合以产生显著性映射(步骤 13)。就 Itti & Koch 模型而言,将显著性数据作为原始数码照片的效果图提供给使用者,其中“最亮的”对象为模型已预测下一个将分配视觉注意的对象。这种所预测的对象在“赢者通吃(winner-take-all)”算法(步骤 15)中被辨认为视觉显著的(步骤 14),并且此过程随后重复,直到多个对象被模型辨认。

[0037] 图 2 是场景 201 的艺术效果图,其可以提供给诸如 Itti & Koch 之类的视觉注意模型。该场景是简化场景,仅为了说明而包含于此;在实施过程中,场景常常是真实数码照片或视频,并且要复杂得多。图 2 包括许多场景内的对象,例如五角星 202、花朵 203、脸 204、五角星 205、箭头 206 和杯子 207。

[0038] 图 3A 是表示 Itti & Koch 模型的输出如何被表示的艺术效果图。突出的(且在该代表性图示中,圈绕的)对象是模型预测将为视觉显著的那些。例如,图中的五角星 202 在突出边界 208 内;花朵 203 在边界 209 内;脸 204 在边界 210 内;五角星 205 在边界 211 内;箭头 206 在边界 212 内;并且杯子 207 在边界 213 内。因而此实例中的模型已确定六个相对于其他对象更为视觉显著的对象。这个特定的模型还预测注意将在确定为在某一视觉显著性阈值之上的对象中如何移动。例如,视觉注意路径 301、302、303、304 和 305 显示所预测的视觉注意路径。

[0039] 图 3B 是示出 Itti & Koch 模型的输出有时被表示的另一方式的第二艺术效果图。除了图 3A 所示的内容以外,图 3B 包括所预测视觉注意的顺序。例如,五角星 202 标为“1”(注意序号 214),花朵 203 标为“2”(注意序号 215)等等。

[0040] 当然,图 3A 和图 3B 仅为视觉注意预测可被传达到使用者的一种方式;不同模型以

不同的方法表示这种信息（或这种信息的某一子集）。例如，虽然可以通过确定具有最高视觉显著性水平的对象，然后排除该对象并寻找下一最高视觉显著性水平的对象等等，来得出这种注意序列，但并非每个模型都确定所预测的注意序列。

[0041] Itti & Koch's 模型代表了一种“由下而上”视觉注意模型，因为该模型基于场景的细节分析进行其预测。其他由下而上视觉显著性模型描述于以下这些参考文献：Gao, Mahadevan 和 Vesconcelos (2008)。

[0042] 除了由下而上模型以外，有另一类模型被称为视觉注意的“由上而下”模型。相比于“由下而上”模型，这些模型以场景和明确任务（例如，避开障碍和收集对象）或将影响注意在特定搜索任务期间将分配在何处的对于世界的现有知识（例如，椅子往往会在地面而不是在天花板）开始。这种知识（既基于任务又基于场景的）与“由下而上”特征结合使用来将注意指向所观察场景内的对象。在 Rothkopf, C. A., Ballard, D. H. & Hayhoe, M. M. (2007) 的“任务和环境确定你看向何处” (Task and context Determine Where You Look) (Journal of Vision 7(14):16,1-20); 以及还在 Torralba, A. 的“目标显著性的环境调制” (Contextual Modulation of Target Saliency) (Adv. in Neural Information Processing Systems 14(NIPS), (2001) MIT Press, 2001) 中描述了一些示例性的“由上而下”模型。例如，视觉注意的 Torralba's 模型具有关于包括特定类型的对象的特征的现有知识以及关于场景内这些对象的绝对及相对位置的信息。这种现有知识对场景内特定目标的搜索提供“由上而下”的影响。

[0043] 本领域已发展为包括既有“由下而上”又有“由上而下”设计特征的混合视觉注意模型，并已适应模型将接触到的场景类型的差异（例如，视频对静止图像，室外图像对网页等等）。

[0044] 稳健性

[0045] 稳健性是指从视觉注意模型输出的预测对以下任一者或其组合的灵敏度：

[0046] (a) 场景视觉特性内的变化和 / 或变动，其包括（例如）场景内对象的布置，对象的照明、对象的颜色等（称为“外部变动”）；或，

[0047] (b) 观察者或观察模型的变化和 / 或变动（称为“内部变动”）。

[0048] 如本文所用的术语“对象”是指场景内的物件、范围或区域，视情况而定，或通过视觉注意模型分析的场景内的区域。术语“对象”视情况而定可与“范围”或“区域”互换使用。

[0049] 两种类型的变化 (a) 和 (b) 被总体称为内部或外部或 IE 变化，引入然后评估这类变化的各种方法在下面进一步讨论。

[0050] 稳健性评估是一种用于测量 IE 变化对注意将分配在场景内何处的影响的方法。然后，场景内对象的稳健性是特定对象的视觉注意模型预测的倾向变化或未变化的程度（不管 IE 变化）。

[0051] 一组对象的稳健性为超过一个对象（一个对象的集合）的视觉注意模型的倾向随着 IE 变化变化或未变化的程度。

[0052] 场景的稳健性是对场景中对象的视觉注意模型的倾向将变化或未变化的程度的测量（不管 IE 变化）。例如，如果视觉注意模型预测将对场景内相同的四个对象进行注意或注视，并且虽然有 IE 变化，但这四个对象往往会仍然被视觉注意模型预测，则该场景往



往会比所述四个对象随着 IE 变化而变化的场景更具稳健性。

#### [0053] 视觉注意稳健性评估系统

[0054] 图 6 是显示视觉注意稳健性评估系统 403 中示例性功能模块的示意图,该系统是用于评估场景的稳健性、场景内对象的稳健性或多个场景内对象的稳健性的系统。当然,也可以手动实践这些方法。在图 6 所示的实施例中,视觉注意稳健性评估系统在计算机系统 408 内。计算机系统 408 可以是任何通用或专用的计算机或装置。其可以是单独的膝上型计算机或多个联网的计算机。另外,计算机系统 408 可以是手持式计算机、数字照相机或平板电脑或甚至移动电话。在一个实施例中,计算机系统 408 具有包括操作系统的各种功能模块(图 6 中未示出)。这样的操作系统便于视觉注意稳健性评估系统进入计算机系统资源。计算机系统 408 具有处理器和存储器以及各种传统的输入/输出接口。

[0055] 视觉注意模块 403 是任何视觉注意模型或模型组合的任何实施例。如前所述,有不同类型的视觉注意模型,但在一定程度上,其全都预测视觉注意往往将会分配到的场景内对象或范围。在图 6 中视觉注意模块 403 示出为视觉注意稳健性评估系统 402 的一部分,但在另一个实施例中,视觉注意模块 403 作为单独的计算机处理或甚至作为在任何类型的计算机网络(例如万维网)上提供在远程计算机的服务而工作。

[0056] VAM 修改模块 404 是视觉注意模块修改模块。VAM 修改模块 404 修改视觉注意模块的参数或架构的一些方面。这种修改可以根据视觉注意模块 403 的执行以许多方法来实现。例如,视觉注意模块 403 本身可以支持对视觉注意模块如何工作的一些方面进行修改的函数调用。在一个实施例中,视觉注意模块 403 可以支持对给予场景某些方面(例如亮度)的权重进行修改的函数调用。在另一个实施例中,如果视觉注意模块经由命令行调用,则可以使用各种开关来改变视觉注意模块内的变量。或者,如果视觉注意模块 403 嵌入脚本或程序代码中,则 IE 修改模块可以修改脚本或程序代码本身。在另一个实施例中,整个视觉注意模型被另一视觉注意模型替换。下面进一步讨论 VAM 修改模块 404 修改视觉注意模块 403 的基础视觉注意模型(或这样的模型于场景的应用)的特定方法,但一个实例可能将修改与用来生成显著性映射的多个特征映射相关的权重。这可以通过将基于可接受值的分布修改这些值的函数调用来实现。

[0057] 场景修改模块 405 修改将提供给视觉注意模块 403 用于分析的场景。场景变化是与模拟环境动态特性的外部环境相关的变化。这样的外部变动可以包括(例如)场景内的移动对象(例如,行人的位置)、由于照明方向变化引起的阴影变化、或者大气条件变化(例如,空气中的粉尘)。这种变动可用许多方法产生。一种方法是将摄影机放在场景内,并在不同时刻捕获图像。这样可捕获真实场景的自然变动。另一方法是捕获场景的单幅图像,并规定对该场景内单个元素和对象进行的可能变化。这样的技术的具体说明如图 7 所示。场景 801 表示的是原始场景。场景 802 示出用场景变动元素替换的场景 801 的对象,所述场景变动元素例如阴影和其他对象(例如,汽车或鸟类等任何可以放在场景内的对象)。场景 803、804、805 和 806 示出原始场景 801 与场景变动元素的不同组合。本领域技术人员会知道有无数种不同的方法来修改场景,例如使用市售的图像编辑软件,例如 Adobe Systems (San Jose, California) 以商品名“Photoshop”市售的软件。本文仅作为非限制性实例示出了几种。

[0058] 场景修改模块 405 将外部变动添加到场景,这又会产生场景或场景实例的多个版

本。在一个实施例中，场景实例的集合留存场景内的变动。为了测量对象的稳健性或场景的稳健性，各场景实例被提交到视觉注意模块 403 以产生关于视觉注意将分配在各个场景实例内何处（也就是说，分配到哪个对象）的预测。从属于各次预测的信息保存在数据库 407 中，然后稳健性评估模块 409 遍及这些不同的实例（并如前文进一步详述）对数据进行评估，以生成基于将被分配模型预测的注意的对象的统计。

[0059] 图形用户接口模块 406 有助于与使用者 401 的交互。图形用户接口模块 406 可以（例如）访问（计算机系统 408 的）操作系统资源来构建图形用户接口以请求用户 401 的输入。在一个实施例中，这种输入包括场景的位置以及视觉注意稳健性评估系统的其他操作参数。在一个实施例中，这种输入将规定评估中用户 401 感兴趣的场景内的区域和 / 或位置。除了规定这类位置以外，用户 401 可以规定何种类型的变动将被视觉注意模块 403 考虑。这可以包括普通或特定内部变动、或普通或特定外部变动。例如，一种特定类型的外部变动可能是由于照明方向的变化而改变图像。随着光线变化，产生的阴影将变化。这是一种特定类型的外部变动，因为其不会考虑如动态对象、大气析光差等的其他因素。普通内部变动的一个实例可以是其中各特征映射的权重被允许独立变化的状况。特定内部变动的一个实例是在一组特征映射的权重（例如，亮度）变化但其他未变化时的情况。图形用户接口模块 406 也有助于（可能从用户）获得有关应该从哪个地方获取场景的图像的输入可能的地方包括（例如）数据库或平面文件。

[0060] 稳健性评估模块 409 控制其他模块的交互，以评估场景内对象的稳健性或场景本身的稳健性。例如，稳健性评估模块 409 必要时调用视觉注意模块 403 以及 VAM 修改模块 404 和场景修改模块 405。稳健性评估模块 409 调用各种模块的方式可以通过来自用户 401 经由（例如）图形用户接口模块 406 提供给稳健性评估模块 409 的输入来修改。稳健性评估模块 409 必要时还评估由其他模块提供的数据并生成报告。

[0061] 数据库 407 处理视觉注意稳健性评估系统 402 的数据存储需求。除了别的以外，数据库 407 可以保留场景的图像。数据库 407 可以是任何计算机存储器。其可以是随机存取存储器或平面文件、或在一个或多个数据库服务器上执行的一个或多个数据库管理系统 (DBMS)。数据库管理系统可以是关系 (RDBMS)、分层 (HDBMS)、多维 (MDBMS)、面向对象 (ODBMS 或 OODBMS) 或对象关系 (ORDBMS) 数据库管理系统。例如，数据库 407 可以是单一关系数据库，例如 Microsoft Corporation 的 SQL 服务器。

[0062] 用户 401 是视觉注意稳健性评估系统的任何用户。在一些实施例中，视觉注意稳健性评估系统 402 非常容易使用，使得不熟悉视觉显著性理论的人员可以使用该系统来评估对象、区域或场景的稳健性。用户 401 可以是评估标牌和非标牌对象在它们的环境内的定位的商业实体雇用的顾问或为其工作的雇员。用户 401 还可以是对评估其中页面的视觉特性可以改变的网页上的数字对象（例如广告）的设计和布局感兴趣的内容设计者。

[0063] 图 8 是视觉注意稳健性评估系统的一个替代实施例，其中视觉注意稳健性评估系统另外包括网络服务器模块 501。为方便起见，网络服务器模块 501 示出为视觉注意稳健性评估系统 402 的一部分。然而，网络服务器模块可以具体化为计算机系统 408 上运行的在单独存储空间中的软件模块。或者网络服务器模块 501 可以在经由网络连接到视觉注意稳健性评估系统 402 的独立计算机系统上。

[0064] Web 服务器模块 501 提供用户 401 可以经其通过客户端计算机 503 并通过网络 502

与视觉注意稳健性评估系统 402 沟通的界面。在一种配置中,web 模块 501 执行网页服务器软件,例如 Microsoft Corporation (Redmond, Washington) 的互联网信息服务器 (Internet Information Server)。网络服务器模块 501 通过使用 (例如) 动态服务器页面 (Active Server Pages)、超文本标记语言 (HTML) 或动态 HTML 编写的网页、Active X 模块、Lotus 脚本、Java 脚本、Java 小程序、分布式组件对象模块 (DCOM) 等等来提供与远程用户 401 交互的机制。

[0065] 尽管举例说明为在由计算机系统 408 提供的操作环境内执行的“服务器端”软件模块,但包括视觉注意稳健性评估系统 402 的功能模块可以容易地实现为在由用户 401 使用的诸如客户端计算机 503 之类的计算设备上执行的“客户端”软件模块。视觉注意稳健性评估系统 402 可以 (例如) 实现为由网页浏览器执行的 Active X 模块,所述网页浏览器在客户端计算机 503 上执行。

[0066] 网络 502 可以是任何类型的网络 (公共网络或专用网络)。在一个实施例中,网络 502 是互联网。

[0067] 图 7 和图 8 中的用户 401 可以是互联网的任何用户。在一个实施例中,用户 401 可以经许可预先安排访问视觉注意稳健性评估系统的功能。

[0068] 图 9 是说明视觉注意稳健性评估系统 402 评估场景内对象或场景本身的稳健性可采用的流程的一个实施例的高级流程图。首先,调用视觉注意模块,并向其提供场景输入 (步骤 601)。然后来自视觉注意模块接收场景内所预测的对象 (步骤 602)。然后,模型预测具有较高相对显著性的位置 / 对象 / 区域的一些指示被存储到数据库 (步骤 603)。确切地说,存储在数据库中的内容很大程度上取决于从步骤 602 中视觉注意模块接收的输出类型。在一个实施例中,数据库存储关于到达阈值显著性值的对象的信息。在另一个实施例中,数据库存储值的矩阵 (图像中辨认的每个对象一个值),并存储 (例如) 对象到达阈值显著性值的次数。然后,进行检查以查看处理是否完成 (步骤 605)。在一个实施例中,这种检查可以确定当前迭代是否已超出由用户 401 最初设置的迭代次数。在另一个实施例中,迭代次数可以由算法确定。在又一个实施例中,迭代次数可以由场景的特性或者由关于稳健性测量的统计数据确定。例如,如果尝试确定在特定场所两副广告中哪副更好,则可以运行该算法,直到两幅广告的稳健性值之间有统计学上可靠的效果。

[0069] 如果稳健性评估未完成 (步骤 605 “否”),则将 IE 变化引入视觉注意模块的视觉注意模型或场景输入。IE 变化可以分为两类:结构化的和随机的。随机变动是不相关的变动。例如,场景中的随机变动可以包括各个像素颜色和 / 或亮度的随机变化。在这种情况下,像素变化是不相关的。相比之下,结构化变动在正被修改的元素之间具有相关性。例如,通过模拟场景内对象的运动或添加或移除场景内对象而修改的场景将构成结构化场景变动。在这种情况下,像素修改的变化是相关的。随机内部变化可以包括被视觉注意模块利用的注意模型中的随机变化。在另一方面,结构化变动可以是注意映射的一部分中相比于另一部分中的注意的程序性偏倚产生变动的方法类型汇总于表 1。引入 IE 变化的子处理在下面进一步详细说明。

[0070]

	内部	外部
结构化	有许多方法来产生内部结构化变动。这种类型的变动由映射或空间区域上的共变限定。一种方法将是改变不同的特征映射对显著性映射的权重。加权值可以因不同试验而略微变化，从而模拟对颜色、取向、亮度等的个人偏好或预期。另一方法是让模型偏置显著性映射的一个区域（例如，下半球），其中所有位置都给予显著性“提高”。	像素集合共变的图像中的变动。这种共变可以通过限定图像内的一个区域（或多个区域）并修改该区域内的 RGB 值而产生。特定区域内的所有像素可被修改至相同的程度。因区域不同或试验不同而变化的量可以通过高斯分布的标准偏差来设置。第二参数可以是可共变区域的“大小”。
随机化	在视觉注意模型确定注意将被分配在何处之前，将随机变化添加到显著性映射的输出。变动可以是任何概率分布的形式，但一个实例中实际使用的是高斯分布，其中高斯型曲线的平均值设置在所期望的显著性值，并且高斯型曲线的宽度由适当的标准偏差来设置。	其中 RGB 值被随机修改的逐像素变化。修改的程度可以由高斯分布的标准偏差来规定，其中高斯型曲线的平均值设置在“真实”RGB 值上。还可以使用其他用于变动的方法。

[0071] 表 1

[0072] 一旦已进行 IE 变化，则处理返回步骤 601。

[0073] 一旦处理完成（步骤 605 “是”），则完成对由视觉注意模块预测的对象在逐次迭代中如何变化和变化的程度的分析（步骤 606）。这种分析在下面进一步描述。

[0074] 最后，生成指示稳健性的输出（步骤 607）。在一个实施例中，这种输出通过图形用户接口模块 406 提供给用户。然而，该输出还可以以其他形式提供到（例如）其他程序或调用函数。

[0075] 视觉注意模型的变化

[0076] 修改视觉注意模型的一个实例（一类可以在图 9 中步骤 604 引入的变化，如上所讨论）是将偏倚程序性地引向场景内的特定特征或场景内的特定范围。给予视觉注意模型的这种变动会对注意将被分配在何处有影响，并将模拟观察者之间的变动和 / 或不同时刻经历场景的人类观察者之间的变动。举例来说，可用 Itti & Koch “由下而上”视觉显著性模型来评估场景的视觉注意稳健性。使用这样的模型，通过合并不同视觉特征映射的输入而生成显著性“映射”。例如，在一些例示中有三幅特征映射：第一幅对颜色敏感，另一幅对取向敏感且第三幅与亮度相关。显著性映射的输入是这三幅映射的加权组合。通常，这三幅映射具有相同的显著性映射权重，指示出一类特征相比另一类特征没有偏倚，并可以表示为权重向量（例如， $[1 \ 1 \ 1]$  表示三幅特征映射具有相等权重）。一种用于生成观察者偏倚的变动的方法是修改这些模拟观察者的权重，所述观察者可能相比另一特征偏向一个特征。例如，可以通过将权重设置为  $[0.5 \ 0.5 \ 2]$  而使模型偏向环境中的较亮物品。一种生成内部变动（或模拟观察者变动）的方法是生成具有不同加权值的模型的集合。这可以通

过随机地设置各模型的权重值来实现。

[0077] 另一方法是随机修改场景的模型内部表达。例如，人类视觉系统中的神经元在一定程度上充满噪声，因为它们的激活甚至会随相同图像的出现而变动。模拟这种变动可以通过扰乱与各个特征映射相关的值的内部表达（例如，颜色、形状、取向、亮度）来完成，所述各个特征映射是视觉注意模型响应场景而形成的。

[0078] 另一方法是扰乱用来辨认图像的哪些区域最显著的显著性映射中的值。例如，可以在每次注视计算后扰乱内部表达，或每当呈现场景时独立地扰乱该值。

#### [0079] 场景的变化

[0080] 基于先前存在的场景（其中场景的特性是各种各样的），通过使用图形编辑器（如，Photoshop）、场景的 3D 表现（例如，虚拟现实模型）或视频（此外，文中全部统称为“场景”）创建新图像来改变场景。例如，这类各种各样的特性可以包括模拟的（或真实的）照明变化或新加或去掉的对象（真实或虚拟的）或随机地改变像素颜色。

[0081] 有许多不同的方法来改变场景。决定进行何种外部变化可以因情况不同而变化。一种用于决定对场景进行何种类型的变化的方法是确定在场景内通常发生的场景变动的类型并将这些变化引到被分析的图像中。例如，在酒店大堂中人群会移动通过场景。有时旅客会站在柜台边，有时站在电梯旁，或从一处向另一处行走。模拟这些情况提供一种用于捕获实际场景中变动的方法。

[0082] 人们可能想要改变和 / 或测量外部变动被添加到场景的程度。一种用于量化外部变动程度的方法是改变场景中被操纵对象的数量。例如，回到有行人的酒店大堂实例，可以改变场景中行人的数量。少数行人将转化成少量变动，而大量行人将转化成大量变动。用于测量变动的另一方法是对生成的全部图像测量像素变化。像素变动的一种实例测量是对生成的一组图像测量各像素的平均方差。一旦算出各像素的变动，就可以通过计算全部像素的均方差来产生单个数值。此外，随着图像内像素的变动增加，该值也将增加。

[0083] 一种用于产生变化的方法是通过使用市售的图像编辑软件（例如 Adobe Systems Inc. (San Jose) 以商品名“Photoshop”市售的软件）数字地修改图像来对图像施加改变。使用这种方法，可以通过数字地放置对象、移除对象或模拟照明变化来修改场景。这些变化随后将被转化成将由模型读取并分析的图像集合。这些变化还可以通过用算法在图像上叠层而自动加到图像上。

[0084] 用于生成外部结构变动的另一方法是生成场景的虚拟再现。使用这种方法，可以方便地修改对象的位置、对象属性和照明条件。可以从这些虚拟再现生成其中对象、其属性、其姿势和照明条件已改变的图像。

[0085] 至此，改变场景的方法可以说是“合成的”，也就是说，其为场景表现的变化，而非场景本身的变化。修改场景的另一方法是修改实际场景。例如，可以拍摄一系列的时延图像（例如，在限定时间内每 300 秒获取一张照片）。这种系列的图片随后可以用于视觉注意稳健性评估系统的逐次迭代。这种分析（跨时间多重图像）将提供外部结构变动的度量。此外，考虑到这种变动（人和对象移动，和由于（例如）太阳位置的变化引起的照明方向的变化一起），可以分析注意通常会被分配到环境内的何处。

[0086] 除了“无分别地”添加外部和内部变动，还可以引入依赖于场景中不同对象视觉显著性的变动。例如，人们可能想要降低注意分配到的区域的变动程度，并增加注意未分配到

的那些区域的变动程度。

#### [0087] 评价稳健性

[0088] 有多种用于表征场景稳健性的方法,例如使用“热图”图形表达、内部熵值、稳健性指数或稳健性值。一种汇总由将视觉注意模型重复应用到场景同时将某一类型的 IE 变化应用于各次迭代所得的综合结果的基本方法,是生成重叠到场景表现上的关联值或区域的图形表达。这样的方法产生类似于热图的东西,其中“更亮的”对象对于所考虑的变动更具有稳健性。

[0089] 除了评价噪声对视觉注意模型预测哪个对象的影响,用于评估稳健性的另一方法是测量场景的模型内部表达的变动。例如,Itti & Koch 模型使用内部“显著性映射”表现来确定视觉注意将被分配的位置。可以测量(但并非唯一方法)随着 IE 变化而变化的显著性映射的内部表达中变化的量或熵,作为稳健性的度量。还可以观察各个特征映射的响应或视觉注意模型的其他内部表达来测量 IE 变化的影响。

[0090] 用于生成稳健性度量的另一方法是 IE 变化可被参数化时的情况。例如,将随机亮度噪声添加到场景(从而通过相对于图像中的亮度能量修改干扰的亮度能量来改变噪声量)就是这样的情况。或者可以增加视觉注意模型的连接权重的变动,并将变化程度表征为熵的度量。用于表征对象、对象组或场景的稳健性的另一方法是确定这些对象在保持其视觉显著性层次的同时可以容许的噪声量。这种度量的方法是测量作为变动量的函数的对象被注意的可能性。在注意开始被吸引到噪声成分而非对象/区域之前,具有高度稳健性的这些对象和/或场景将容许大量的噪声。所得的度量可以是在注意从对象引开并引到噪声之前对象可以容许的噪声水平。

[0091] 图 4 是示出注意对象的概率与两个对象的变动程度的关系的曲线图,第一对象 451 具有低噪声容限(低稳健性值),并且第二对象 452 具有高噪声容限(高稳健性值)。曲线图显示变动对注意是否会被分配到对象 451 和对象 452 的影响,每个对象具有各自不同的稳健性指数。在此实例中,添加少量的变动显著地影响模型可预测注意被分配到对象 451 的概率。这由图 4 中在仅有少量变动时的概率降低证明。相比之下,对象 452 未被少量的变动影响。这由随着更大变动等级出现的“下降”点说明。通过测量这些“下降”点,可以用稳健性指数来量化场景内的不同对象。

[0092] 除了刚描述的热图方法,有时量化场景内特定对象(或多个对象)的稳健性也会是有用的。这种量化可以便于后续的分析,例如确定开始减少场景内特定对象显著性程度的 IE 变化的程度(或稳健性指数值)。

[0093] 有时评估和/或量化场景本身的稳健性也会是有用的。对象和场景稳健性之间的这种差别使许多种令人感兴趣的后续分析成为可能。例如,场景可能具有低稳健性值(也就是说,模型预测将受到视觉注意的对象组随着 IE 变化而变化),而该场景内的一个对象具有高稳健性值。“高”对象稳健性值和“低”场景稳健性值的例子将是(例如)在目标对象总是处于前三位注意注视但其他对象被视为显著变化时。

[0094] 表 2 提供示例性数据,其中具有高和低稳健性值的目标对象与具有低和高稳健性值的场景进行排列组合。在此表中,示出了一种可能的稳健性度量,其中,为每一个对象计算归一化的似然值(下面称为稳健性值)。对于各个场景而言,左列表明当有 IE 变化时对象在前三次扫视中出现的时间百分比。在表 2 所示的实例中,场景内有 14 个对象,一个目标

对象和 13 个干扰物,干扰物是具有显著性的非目标对象或区域。如果模型要随机选择三个对象,则各个对象可具有 21.42% 被选中的概率 ( $3 \times 1/14 = 0.2142$ )。这可以是其中极少有对象和 / 或场景稳健性的情况,并因此其用作重要基线,以通过该基线比较场景的稳健性。这是因为稳健性随着概率接近机遇概率而下降。在此实例中,稳健性值在数学含义上是被注意百分比除以被选机遇百分比,从而 4.6667 的第一稳健性值 =  $100\% / 21.42\%$ 。一种用于计算场景稳健性值的方法是计算前 K 位对象 (注意选择的次数) 的平均稳健性值。表 2 的底部是前 3 位对象的平均稳健性值。

[0095]

稳健性	目标对象=高 场景=低		目标对象=高 场景=高		目标对象=低 场景=高		目标对象=低 场景=低	
	被注意的	稳健性值	被注意的	稳健性值	被注意的	稳健性值	被注意的	稳健性值
目标对象	100%	4.6667	100%	4.6667	5%	0.2333	9%	0.4200
干扰物_1	15%	0.7000	90%	4.2000	97%	4.5267	30%	1.4000
干扰物_2	15%	0.7000	5%	0.2333	5%	0.2333	26%	1.2133
干扰物_3	17%	0.7933	2%	0.0933	2%	0.0933	15%	0.7000
干扰物_4	14%	0.6533	80%	3.7333	88%	4.1067	32%	1.4933
干扰物_5	13%	0.6067	3%	0.1400	91%	4.2467	28%	1.3067
干扰物_6	21%	0.9800	2%	0.0933	2%	0.0933	22%	1.0267
干扰物_7	18%	0.8400	1%	0.0467	1%	0.0467	32%	1.4933
干扰物_8	19%	0.8867	4%	0.1867	4%	0.1867	33%	1.5400
干扰物_9	16%	0.7467	2%	0.0933	2%	0.0933	38%	1.7733
干扰物_10	20%	0.9333	1%	0.0467	1%	0.0467	18%	0.8400
干扰物_11	14%	0.6533	2%	0.0933	2%	0.0933	17%	0.7933
干扰物_12	18%	0.8400	8%	0.3733	8%	0.3733	9%	0.4200
干扰物_13	19%	0.8867	0%	0.0000	0%	0.0000	30%	1.4000
场景稳健性值	2.1933		4.2000		4.2933		1.6022	

[0096]

[0097] 表 2

[0098] 表 2 按对象对注意被分配的位置作了分类。但如上所述,术语“对象”被宽松地定义为图像 (或场景) 中的区域或范围。然而,本文所述的方法和系统并不限于基于对象的方法,对其他类似的方法也将有效。例如,等大区域的网格可以在场景或者基于人类视觉系统特性 (例如,观察者的中央窝的大小) 限定的场景区域上定义。

[0099] 评估跨场景稳健性

[0100] 考虑到上述用于评估对象或场景稳健性的方法,以其他方法扩展稳健性的评估是近乎可能的。例如,“跨场景稳健性”是对象 (或对象组) 的稳健性如何随不同场景而变化

的度量。通过为与场景分开的对象提供稳健性度量（如上面所讨论），可以评估不同场景中特定对象的稳健性。例如，考虑广告商在将放在三个不同场景中三个不同广告牌上的两段不同的广告内容之间进行决定的情况。图 5 提供这种情况的实例的具体说明。场景 703 包括广告牌 700。广告牌 700 在场景 704 和 705 中重复。类似地，广告牌 706 见于场景 707、708 和 709。

[0101] 使用前述的方法（例如将广告内容叠加在场景的图形表达中），可以确定在三个场景的每一个中各个广告牌的稳健性值，从而生成六个不同的对象稳健性值。客户随后可以选择具有最高平均稳健性值（或基于稳健性值集合的其他度量）的广告（对象）。

[0102] 挡手头有一组对象但需要从一组场景中选择将一个对象或多个对象放在哪个场景中时，也可以使用类似的分析。继续讨论该广告牌广告实例，客户可能有他们想要置入三个不同场景中的两个的单幅广告。可以使用上文描述的方法来生成数字图像，所述数字图像可以用来为三个不同场景中的各幅广告生成稳健性值。为便于场景选择，客户可以选择具有最高稳健性值的两个场景（广告叠加在场景内）。

[0103] 除了为固定场景组从对象集合中选择单个对象（例如，广告牌）或为单个对象（从场景集合）选择场景组，还可以基于稳健性值将对象分配到特定场景。例如，客户可能有其已开发的三幅不同广告（放在广告牌上的内容）和二十个不同的可放置这些标志的广告牌。通过为放在二十个场景中的三幅广告每个生成稳健性值，客户随后可以选择将具有最高稳健性值的广告用于各个单独场景。另外，可以将最佳广告分配到具有最高稳健性值的十个场景。

[0104] 虽然这些实例至此已说明了稳健性度量可以如何用来改进关于广告牌和广告进行的选择，但稳健性度量的有益效果不限于这个领域，并可以用于其中有背景（场景）的集合和视觉刺激物（对象）的集合可在其间选择的任何领域，例如商店或酒店内的数字标牌、静态标牌、产品标牌、产品包装配置或网站。

[0105] 应当注意，本文所述的用于评估场景内对象稳健性或场景本身稳健性的系统和方法不依赖于用于确定视觉注意的任何特定方法论。相反，这些系统和方法通常可以与用于评估视觉注意的任何模型一起使用，并且在一些实施例中，可用多个不同视觉注意模型来评估对象或场景的稳健性。

#### [0106] 稳健性和注意注视顺序

[0107] 到现在为止的讨论主要集中在场景中对象的稳健性或者该场景的稳健性或该场景内对象集合的稳健性，而没有考虑对象被注意的顺序。例如，到现在为止的分析仅对模型是否实际上预测注意是否被分配到对象作了评估。然而，在一些情况下元素的顺序实际上确实很重要。例如，对于多对象场景而言，人们可能想要知道一个对象在另一对象之前出现的频率以及在对两个单独对象的注意之间是否有被注意的介于其间的对象（所谓的居间对象）。

[0108] 利用上文描述的涉及将变动引到场景的方法，人们会获得注意注视顺序的集合。使用这个数据，人们可以辨认其中实现了目标顺序（例如，在对象 B 之前注意对象 A）的所有序列。顺序稳健性分析可以将实现目标序列的概率用作序列目标稳健性的度量。一个用于测量序列目标稳健性的实例方法是计算（如果将要随机选择对象）目标顺序（也就是说，对象 A 在对象 B 之前）会发生的可能性。序列目标稳健性值可以是实现目标序列的概



率除以其偶然发生的可能性。这种度量背后的原理类似于上述稳健性值背后的原理。

[0109] 考虑第一种情形,其中注意的相对顺序确实重要但介入其间的对象并不重要。目的是让注意在对象 B 之前被吸引到对象 A,例如公司在靠近餐厅本身的广告牌上发布餐厅晚餐特色菜的广告将会是这种情况。公司会考虑两幅考虑周到的晚餐特色菜广告。目的是让经过的人们首先注意晚餐特色菜广告,然后才注意餐厅本身。使用上面所述的方法来评价这两幅不同广告(也就是说,将广告数字地插入数字图像,从而使其看上去就像被定位在它们被放入场景内的广告牌上),然后施加 IE 变化,就可以算出广告牌在餐厅之前受到视觉注意的频度。表 3 和表 4 提供可能的似然性,表示对象 A 和对象 B 的相对排序以及它们是否受到任何注意(对象 -# 未被注意)。从这些表我们看出在广告内容 1 的情况下对象 A 在对象 B 之前出现 65%的时间,但在广告内容 2 的情况下仅 40%。从而,如果这个顺序很重要,则客户会倾向于选择广告内容 1。

[0110] 广告内容 1

[0111]

	对象 A 位置 1	对象 B 位置 2	对象 A 未被注意
对象 B 位置 1		25%	7%
2.对象 B 位置 2	65%		
对象 B 未被注意	2%		1%

[0112] 表 3

[0113] 广告内容 2

[0114]

	对象 A 位置 1	对象 B 位置 2	对象 A 未被注意
对象 B 位置 1		20%	20%
2.对象 B 位置 2	40%		
对象 B 未被注意	5%		15%

[0115] 表 4

[0116] 稳健性和多样本条件

[0117] 至此的讨论集中从单一位置观察单个对象的情况。然而,世界是高度动态的,并且很多时候观察者正在穿越空间移动。当观察者穿越空间移动时,观察者可以有多次机会来“处理”一个特定对象。然而,这种移动产生稳健性分析的另一重要方面,这个重要方面就是其中对象可见的时间或者注视次数可能变化。我们将这个称为可见持续时间-观察者可看见一个特定场景或对象的时间长短。一种用于捕获可见持续时间的方法是通过使用报告视觉注意位置的预测顺序的模型。通过这些模型,可见持续时间可以通过限制对于场景的稳健性评价模型所认为的注视次数来规定(这也适用于下文讨论的场景优化)。较长可见持续时间与更多的注视对应,而较短可见持续时间将对应于较少的注视。可见持续时间的变化可以对稳健性分析具有显著的影响。

[0118] 当处理场景的注视次数受限制时,对象是否会受到注意将取决于其在视觉注意序列中的位置。例如,客户可以将数字标牌放在酒店内。数字标牌呈现两个内容块-一个在

另一个之后。一个内容块呈现 3 秒,第二个呈现 6 秒。考虑到其有限的可见持续时间,3 秒长的内容块需要在注意序列中比 6 秒长的内容块更早地出现。例如,如果人们每秒进行 2 次注视,则模型须预测到在前六次注视中对象会受到注意。相比之下,6 秒内容块必须在前 12 次注视中受到视觉注意。顾及这些动态和像它们一样的其他动态,不考虑可见持续时间会导致关于人们在场景中将注意对象的不准确预测。

[0119] 为了说明这点,我们将拓展该广告牌实例(但这种相同的构思适用于动态变化的任何显示器)。让我们设想一条很长的直路,其中沿着该道路有多块标牌。此外,该道路是平坦的且除了广告牌之外没有其他对象。在这些条件下,所有广告牌将具有相同的可见持续时间。也就是说,视觉系统处理任何特定广告牌的时间量是相同的,并将由汽车的速度、标牌的尺寸和场景中任何大气析光差来确定。

[0120] 让我们考虑这条相同的道路,其中在道路上的最后一个广告牌之前 500 英尺处有一排树木。其他所有广告牌将具有相等的可见持续时间,但这最后的广告牌将具有较短的可见持续时间。考虑到这种可见持续时间的变化,人们会想要限制模型考虑用于稳健性评价的注视次数。第一组广告牌可以具有高的注视次数,而当考虑最后广告牌的持续时间时,人们会考虑影片的较少画面或图像中较少注视。

[0121] 当考虑场景内有多个对象的情况时,可见持续时间这方面可发挥重要作用。在一些情况下,人们可能想要将注意捕获元素分布得足够显著以被处理(即捕获注意),但不是如此显著以致其分散对场景内其他对象的注意。(注意:此概念还关系到场景优化,场景优化涉及以实现特定目标的方式修改场景。场景优化在下面进一步讨论。)鉴于此,当分析对象、对象集合或场景的稳健性时,人们会想要考虑可见持续时间。例如,对于在路径上具有较短可见持续时间的对象而言,人们可能想要增加显著性元素(例如,运动、亮度、色彩对比度等)以增加在该较短可见持续时间期间一个特定对象会被处理的可能性。然而,对于沿着该相同路径可见、但具有较长可见持续时间的另一对象而言,利用有多次机会来处理该对象这一事实,人们可以将较少的(或较弱的)显著性元素分配到该对象。

[0122] 考虑到对象是否沿着路径在任何地方受到视觉注意,可见持续时间可成为稳健性分析的因素。从而比其他对象具有更长可见持续时间的特定对象可获得与之相关的更多注意注视,并且如果注意被分配到在沿着该顺序的任何位置处的对象,则其将被视为“命中”。因此,具有较长可见持续时间的对象将具有更多样本,并将具有更高的被处理的可能性,于是可要求较低的待处理显著性。具有较短可见持续时间的对象将具有较少样本,从而可能被较少地注意,于是可能需要较高的显著性元素以在该较短的序列期间被检测到。

#### [0123] 场景优化

[0124] 至此,本公开的内容集中在稳健性上。现在我们转向与其他视觉注意建模相关的构思,在一些实施例中,其可以得益于前文描述的与稳健性相关的方法和系统,但不一定需要稳健性评估。一个与此相关的概念是场景优化。

[0125] 如较早前所讨论,现有的视觉注意模型可预测受试者将其视觉注意分配在场景内何处。然而,这类模型不提供任何用于指定场景可以如何修改以实现特定视觉目标的机制。因为人类视觉系统实际上不会处理场景的全部可视区域,但相反只处理注意被吸引的那些区域,在许多真实世界情况下,最好不仅让人们“观察”场景内的特定对象,还让人们“注意到特定对象”。

[0126] 视觉目标因此指的是受试者能注意到场景内对象的想望方式。例如,除了视为不重要或甚至不利的对象集合之外,视觉目标可以仅仅是特定对象(也就是说,人们(根据视觉注意角度)确定为重要的场景内的对象集合)被注意的愿望。或者,视觉目标可以是让特定对象以特定顺序或在特定时间被注意的愿望,或者是让特定对象从特定视点但不必从其他视点被注意的愿望。本节的剩余部分讨论可利用视觉注意计算模型来优化场景以实现视觉目标的方法。将要讨论的方法中有一些使用稳健性评估,如前所述,但有些方法并未使用。

[0127] 参见图 10,一旦已限定视觉目标(步骤 1201),就通过将显性成本/回报结构分配在场景内的对象上来开始场景优化(步骤 1205)。这样的分配可定量地确定视觉目标。成本/回报结构限定场景内哪些对象是高值对象、低值对象和甚至用户将其视为将视觉目标分散或对视觉目标不利的对象。用户会将“注意效用”值设置到被认为是视觉目标的一部分(正回报)或不利(负成本)的各个对象上。或者,用户可以设置指定哪些元素比其他对视觉目标“更”有价值的优先级。

[0128] 其次,限定场景的多个可能变化(步骤 1210)。这些可以是简单的变化,例如照明、对象的颜色、对象的定位等,或更复杂的设计变化,例如接待区应当建在大堂内何处。当然,理想的是,与定位接待区同样重要的事情的评估时间最好定在酒店大堂建造之前,从而可期待本文所述场景优化方法的一个效用将用于评价这类场景内的合成或部分合成的场景及设计/布局选择。

[0129] 然而,实际的真实世界“成本”(或估价)与各个可能的场景变换有关(步骤 1215)。例如,在酒店大堂的情况下,目标是让老顾客注意到特定标牌,改变标牌的颜色会是相对低成本的(可以确定 \$200 的估计改变价),而改变花岗岩地板的颜色会是高成本的(可以确定 \$20,000 的估计改变价)。确定真实世界成本估价使得可以将价格数字与多个变化相关联。例如,可以考虑场景的可能变化范围,一些变化满足所有目标,一些变化满足大部分目标,一些变化满足目标并且以最低成本实现,其他变化满足 90% 的限定目标,并为实现剩余 10% 的目标,会发现要花费不成比例的投资资本。换句话讲,在一些实施例中,将真实世界成本与可能的变化相关联便于更有效地对选择方案作出估价。最终,该方法可提供将注意效用最大化而同时将与对象特征配置相关的成本最小化的场景配置(步骤 1220)。

[0130] 应当注意,真实世界成本仅为特定变化的相对成本可以如何被共同相关的一个例子 - 也可以同样容易地使用其他系统,例如较高点与特定变化的更高成本相关且较低点为成本较低变化的点数系统。

[0131] 具有跟踪并调控不连续对象的能力的图形编辑器可用于给场景限定可能的变化。例如,可以辨认在这样的图形编辑器内视见的场景中的标牌(例如,右击鼠标),且用户能够选择该对象的可变特性。这些可变特性可以包括颜色、照明、层内设置(例如,对象可以放在其他一些对象前面或后面)、对比度、阴影、尺寸等。除了选择可以改变的单独特性以外,用户还可以有能力来限定变化或其他相关参数的允许范围。例如,就颜色而言,客户或用户会发现适用于场景内特定壁面的唯一颜色可以是棕褐色的色调。因而颜色属性被限定为仅在规定光谱范围内变化。类似地,如果属性是尺寸,则在对象的尺寸为客户或用户所接受之前,会有对特定对象可以增长到(或减少到)的尺寸的明显限制。如前所述,真实世界的成本估计与各个可能的变化相关。如果有可能变化的范围,则用户可以限定在该范围内

的示例点的成本,并且支持计算机系统(后面所述)将外推出这些实例点的最佳拟合曲线。会提供给用户多个可能的最佳拟合算法,并因而要求用户选择使用一种算法。例如,用户可能只想要在标牌的尺寸方面限定最小容许标牌的成本为 \$100,并且中等标牌的成本在这两个成本点之间线性(基于尺寸)增加。

[0132] 随着目标(或多个目标)被限定、目标的注意效用、可能的场景变化和可能场景变化的成本被限定,下一步骤是评估可能场景变化的有益效果,并尝试寻找实现步骤 1201 中所限定视觉目标的场景配置。在一个实施例中,这通过在算法上修改场景的特性以最大化预期回报、或最小化成本而最大化回报的场景配置、或仅仅满足为最低成本所限定的目标的场景配置来完成。预期回报作为特定对象将受到注意的可能性和这些对象被注意的回报/成本算出。除了注意场景中位置的回报/成本之外,优化过程也将对场景进行特定类型的变化(例如,改变场景内对象的颜色和/或位置)的成本/回报列为重要因素。可以使用前节所述的稳健性分析来确定一种用于计算对象受到注意的可能性的方法。

[0133] 式 1 提供如何使用视觉注意的计算模型来公式化效用函数的一个实例。

$$[0134] \quad ER(F) = \sum_{a_f \in A_f} [a_f(xy)R_a(O_f(xy))] + \sum_{o \in O} R(o_f)$$

[0135] 式 1

[0136] F 是场景内对象的可变属性的集。例如,其将包括场景内各个对象的颜色、纹理或位置。A 是在该特征配置 F 下模型预测的注意注视的集。R(O<sub>xy</sub>) 是注意到位置 (xy) 处的具有特征集 f 的对象的回报(正和负)。a(xy) 是模型预测注意将被分配到位置 xy 的可能性。对于视觉注意的一些模型和方法而言,(xy) 可以是二进制值(0 或 1 指示注意会被或不会被分配到该位置),而对于其他模型和方法而言,这可能是注意将被分配到该位置的可能性(0...1)。无论如何,公式的这部分规定了对当前对象特征集的注意分配的回报。

[0137] 函数的其他部分规定了使用对象 o 的特征集 f 的成本。R(O<sub>f</sub>) 规定了在对象 o 上使用特征集 f 的回报/成本。在一些情况下,特定特征的成本可以是 0.0(例如,对象 o 的当前特征集可能会是这种情况)。然而,人们可能想要模型将所有可能的特征组合看作是同样困难的。在这种情况下,所有特征的回报将是相等的(或最容易地为 0.0)。然而,在一些情况下(例如,改变场景中地毯的颜色与移动花瓶比较),将会有对容许特征配置的限制。在这类情况下,人们可以在对象(o)和特征(f)的成本/回报方面规定这些值,或者其可以仅仅避免将非容许特征配置限定为容许变化。

[0138] 使用该回报函数,探究优化所述回报函数的特征配置的解空间。一旦指定了回报函数,就有多种方法用于求出最优解。这些方法包括(但不限于)使用封闭方程、蒙特卡罗模拟法、模拟退火算法、遗传算法和随机梯度下降法。除了这些近似方法以外,对于一些视觉注意模型而言,可以进行封闭式分析。

[0139] 然后可以得到满足目标的解空间的解以及相关的成本信息以作评估。

[0140] 应当注意,至此许多实例集中在通过增加场景内对象的视觉显著性来优化场景。然而,应当注意,一些视觉目标可以通过降低来自其他对象的视觉注意(或分散)而实现。根据成本模型如何建立,本文所述的场景优化方法可以在场景的噪声抑制方面取得结果(不总是让对象更为视觉上显著)。

[0141] 场景优化和注意序列

[0142] 如上所述,一些目标可不考虑对象受到注意的顺序。在这类条件下,注意到对象的回报/成本将不受其在顺序中的位置和/或在当前注意注视之前或之后受到注意的对象的影响。然而,存在其中场景内对象的注意注视顺序可能是重要的情况。例如,当受试者遵循特定指令序列时,顺序往往有用。

[0143] 为了处理这种更复杂的视觉目标,上述预期回报函数(式 1)可被扩展来优化特征配置,以使预期回报依赖于顺序。这可以通过利用随着扫视次数变化而变化的可变回报结构来实现。应当注意,在式 1 中注意到对象的回报由单独注意扫视( $R_a$ )指示。通过基于扫视的顺序位置( $a$ )指定不同的回报,可以产生由所预测的扫视序列优化场景的方法。图 11 是说明两个回报结构的曲线图。一种回报结构由不随着扫视位置而变化的量表示(线 1225),第二回报结构依赖于扫视位置(线 1230)。依赖于扫视的回报结构规定如果注意到该特定对象在早期发生则其预期回报非常高,但随着注意后来在顺序中被分配而降低。这种类型的回报结构可能与“高值”对象(例如建筑物区域场景中的行人)相关。

[0144] 还可以不将回报建立在对象多早在注意序列中被注意的基础上(如图 11 所示),而是将回报建立在基于顺序的目标的基础上,其中对象的回报基于已在其之前和在其之后受到注意的那些对象。例如,酒店业主可以有两个户外标牌。一个为其餐厅的特色菜做广告,第二个显示其酒店的名字和标志。酒店业主认定,重要的是广告特色菜标牌应当在酒店标牌之前被看到。考虑到该视觉目标,分析会将非常高的回报放在“餐厅特色菜”标牌在酒店名字之前受到注意的情况。而且,对于酒店名字在“特色菜广告”标牌之前受到注意的情况给予低回报,如果其中任一个被注意到而另一个未被注意到,则可不给予回报。顺序可以是相对顺序(之前与之后比较)或者可以是绝对顺序(除非对象 A 作为受到视觉注意的第一对象出现,且对象 B 为受到视觉注意的第二对象,否则对于注意到对象 A 和对象 B 无回报)。当然,正如本领域技术人员会理解的那样,可以有很多其他的制订回报结构的方法。

[0145] 至此,本节已讨论了两个可能的回报函数,在这些函数中扫视的位置或扫视的顺序影响回报。本领域技术人员将认识到人们可以在对象集和扫视集上定义任何的任意顺序回报函数。更一般地说,人们可以通过定义具有讨论中扫视序列长度的 M 维空间(每个对象一个维度)来定义回报结构的这个顺序分量。

[0146] 给定了回报结构就可建立优化过程来为场景提供最佳配置,但会发现所得到的场景“建议”并不合适。例如,人们可能想知道如果修改一个对象的颜色以使其更显眼或更不显眼,注意序列将会如何变化。使一个对象更显眼会对所预测的顺序有非显见的非线性影响。这是因为注意分配就其定义而言是一种零和博弈这一事实:将注意分配到一个对象将必然意味着注意从另一对象离开。因而修改一个对象的特征来增加其明显性将不仅改变该对象会受到注意的可能性(以及其将在注意序列中何处受到注意)而且还将影响其他对象会受到注意的可能性以及那些对象在注意序列中何处受到注意。

[0147] 因为达成优化过程的方法之一是自动考虑多个特征配置,将探究定义解空间的多个(或全部)可能配置。用户或客户会对非最优解极感兴趣。例如,人们可能想要知道什么颜色让标牌将其位置从比如说扫视序列中的第 25 位移动到前 5 位。系统会快速查看所存储的其中所有其他对象的特征保持恒定且感兴趣的对象落入前 5 位扫视的注意序列。

[0148] 如前所述,稳健性还可以用于场景优化。例如,人们可以不仅在该图像的特性方面优化场景,还可以在一天内的不同时间可能出现的场景方面或在不同观察者方面优化场景。

也就是说,优化过程可推荐对场景可能经历的变动具有稳健性的场景。此前我们描述了用于捕获和生成对于输入场景和模型的内部和外部变动的方法。这些方法的目标是模拟(或捕获)实际场景的预期变化。例如,一种用于捕获观察者变动的方法是改变视觉注意模型的参数(例如,用于计算显著性的单独特征映射的权重)并在相同图像上运行这些模型。通过不同模型参数(内部变动)的多个具体实例运行各个场景配置,将给予特定场景配置多个评分-每个模型一个注意序列。通过获得每个模型配置的平均分数,可以生成具有给定变动的场景配置的预期得分。给出最佳平均分数场景配置可得到推荐。

[0149] 另外,可以按照稳健性来限定视觉目标,然后将场景的一个对象(或多个对象)优化到特定的稳健性值。

#### [0150] 可见持续时间

[0151] 如上面就稳健性讨论方面所述,不同的对象或不同的场景可能会有不同的可见持续时间。提示一下,可见持续时间是指对象和/或场景能被看到的时间段。以时间规定的可见持续时间通常将被转换成将在优化过程中考虑的所预测注意注视(或显著区域)的数量。可见持续时间可用来限制会被用于各种场景分析的注视集。更具体地讲,其将被用于视觉目标分析。

[0152] 当考虑多个视觉目标时,可见持续时间会有影响。举例来说,人们可能有具有六个视觉目标的文件或内容,所述六个视觉目标具有不同回报,比如说回报值为1,2,3,4,5,6。如果内容在数字标牌上显示5秒,且人们每秒进行约2次注视,则其可转换成10次注视的可见持续时间。假如可见持续时间是10次注视,则可见持续时间长到足以捕获所有视觉目标。在这种情况下,模型将使六个项目的显著性相对相等(假设场景中没有其他对象且作出变化的成本相等)。使用可见持续时间的优化过程能使视觉目标大致相等,其原因是模型在这种情况下更可能注意到所有对象。如果目标之一(比如说具有最高回报的对象)的显著性明显高于其他对象之一(比如说具有最低显著性的对象)的显著性,则大部分视觉注意模型将首先注意到最显著的对象,然后注意到第二最显著的对象,但最终通常会再返回到最显著的对象。如果返回到前一对象没有额外的回报,则这种视觉注视将不会增加总体回报。然而,通过使对象显著性相对地等同,模型更可能注意到所有目标对象,并因而在给定的可见持续时间下实现更多(如果不是全部)的视觉目标。优化过程和可见持续时间的这个方面非常不同于仅仅使对象的显著性与目标对象的回报或相关物相关联。

[0153] 图12是场景优化系统1255的高层次图。该图中类似命名的元件在功能和能力上类同于前面讨论的那些模块。场景优化系统1255包括数据库图形用户接口1245,其通过网络502经由计算机503接收用户401的输入,以限定场景的一个或多个视觉目标以及可能的场景变化。场景中对象的这些可能变化存储在数据库1235中。场景优化模块1260反复探究实现视觉目标的可能场景的范围,并调用场景修改模块1240来以与由用户401限定的可能变化一致的方式修改场景输入。这产生了被修改的场景,其被提供到视觉注意模块403,视觉注意模块提供与视觉注意有关的输出,该输出存储在数据库1235中。如果用户401在远端(并如图12所示),则场景优化系统1255可以包括网络服务器模块1250。

#### [0154] 多视角场景分析

[0155] 至此,描述主要集中在在使用单幅静态图像或影片来预测注意将分配在一副图像或多幅图像内何处的视觉注意建模。这两种方法常用于许多情况,然而,其局限性在于将复杂

的三维场景中单个二维视图上操作。

[0156] 在我们生活的 3D 世界中,视角(取向和/或位置)的小变化会对投射在观察者视网膜上的图像有显著改变。从一个视角可见的对象可能在另一视角完全不可见。此外,对象之间的空间关系(也就是说,一个对象相对于第二对象在视网膜上投射图像的位置)可以因视角的不同而显著变化。因为视角的变化可以对投射到视网膜的图像产生较大变化,所以其也可能对人类视觉注意被分配到场景内何处具有显著影响。例如,图 13A 是包括在服务台后面具有数字标牌 1290 的服务台的酒店大堂场景的艺术效果图。场景已通过视觉注意模型进行了分析,所述模型预测图 13A 中吸引最多注意的两个对象是数字标牌 1290 和广告卡片 1285。图 13B 是相同酒店大堂中的相同接待区,但数字标牌 1290 不在由视觉注意模型辨认的若干对象当中。如果视觉目标是让老顾客注意到数字标牌 1290 且只有图 13A 中所示的场景被分析,则可能会有视觉目标始终被满足的虚假安全感。因而需要有一种有效的方法来测量和评估 3D 场景内的场景内容。

[0157] 应当注意,3D 场景评估(在本文我们将其称为多视角场景分析)与仅仅将 2D 视觉注意分析延伸到诸如电影序列之类的事物不同。这样的方法可以提供来自多个视角的数据,但最终具有类似于单幅图像方法的缺陷,原因是其可以为人们提供分析通过空间的特定运动的注意分配的能力,但其可能不提供考虑通过空间的其他路径和/或运动的能力。不论是静态场景还是连续的多个静态场景(视频)的视觉注意分析都不能有效地处理图像源自 3D 环境这一事实。

[0158] 因此,拥有适应三维空间和观察者可以穿越三维地理空间的种种手段的多视角视觉注意建模方法是有效的。

[0159] 图 15 是示出多视角视觉注意建模方法的流程图。该方法从总平面设计(步骤 1329)开始,总平面设计包括确定 3D 场景内想要分析的位置和对象。实践中,这可以意指获得或开发将被分析的 3D 场景的平面图,并确定平面图中将代表用户观察视角的位置。当然,在没有 3D 场景的平面图的情况下,可以使用一种较不严格的方法,其中使用者仅仅前往现场并确定哪些位置是感兴趣的,并从这些位置拍摄照片。另外,用户可以记录图像拍摄地点的位置和取向,其可用于记述。

[0160] 除了照片以外,还可以使用其中画面从视频或视频顺序捕获的视频技术。当使用画面(来自视频的图像)或图像(例如,来自数字照相机)时,用户还可以使用视图插补技术用来产生两幅不同图像之间的视图。如前所述,图像不限于由真实环境产生,但其也可以由合成(虚拟)环境产生。然而,在这两种情况下,必须记录或预先指定图像拍摄地点的位置以及摄影机的视角(摄影机指向哪条路线)。一种简单的方法是指定位置,然后让各个连续图像从由围绕北、东、南、西轴线前进 90 度的视角产生(如下面实例所完成的)。但在没有预先指定位置和摄影机瞄准方案的情况下,可以使用 GPS 跟踪技术与光学跟踪技术相结合来代替摄影机进行跟踪。例如,一个器械或多个器械将附接到摄影机,以便每当获取图像时,系统将记录三个位置维度(X,Y,Z)以及三个取向维度(摇摆、倾斜和滚转)以明确表示捕获图像的观察点。这六个值将存储在存储器中,并与在当时捕获的图像关联。当然,可以手动记录摄影机位置和取向。

[0161] 随着 3D 场景被确定,然后接收 3D 环境的多幅图像,该多幅图像表示随着观察者通过 3D 环境进行交互和导航(步骤 1330) 观察者会经历的视图的集合。在一个实施例中,这

通过获取多幅照片来实现,所述照片从 3D 场景内(如仍为酒店大堂)的多个位置和取向拍摄。在另一个实施例中,视频由 3D 场景制成,从预计观察者可能出现的多个代表性区域拍摄照片。在又一个实施例中,使用虚拟 3D 模型,并通过将虚拟摄影机移动通过虚拟 3D 模型来生成视图。不管如何生成,结果是多幅来自 3D 场景内的各种位置的 2D 图像。如果未预先指定,则也可以收集代表 3D 环境内位置和摄影机取向的数据。该数据将允许人们从多个视角评估场景,同时评估观察者可能选取通过 3D 空间的多条不同路径。

[0162] 一旦收集到图像,则选择并辨认对应于所感兴趣对象的二维图像的区域(步骤 1335)。这可以使用多个不同的方法来实现,所述方法可以包括(但不限于)自动提取这些对象的方法、对象的人工选择或甚至是使用自动和人工的标记和标签技术的混合方法。一种用于实现这个过程的方法的图形结果的具体说明示于图 14A 和图 14B。诸如壁画 1310、数字标牌 1315 和花瓶 1320 之类的对象已全部通过人工选择过程加以辨认。

[0163] 其次,用视觉注意模型处理包括感兴趣对象的图像以产生视觉注意数据(步骤 1340)。如前所述,一个这样的模型是 Itti 和 Koch(2001) 所描述的模型,但任何视觉注意模型均可使用。随后将观察者处于各个观察点时模型预测视觉注意被分配的二维位置记录在(例如)数据库中。这些位置随后与已为各个观察点标记和标签的区域比较以确定模型预测可视区域内哪些对象将受到注意。

[0164] 在为各个单独观察点分析各个单独观察点并计算模型预测哪些对象将吸引视觉注意后,生成并存储关于哪些对象将从哪些位置被注意的数据。该数据随后可以用来(例如)确定随着潜在观察者穿越 3D 场景特定对象将被其(完全)观察到的可能性;在 3D 场景内事实上观察到特定对象的潜在观察者的百分比;为特定视图的子集确定对象受到注意的可能性(可用于分析关于特定移动路径的信息,例如进入大堂与离开大堂比较)、当对象看得见时该对象将受到注意的可能性(一些对象可能需要从环境内大部分的可能观察点看上去不显眼,但就对象看得见的观察点而言,人们可能想要得到注意被吸引至该对象的高度确定性)、或 3D 场景内可见到对象(或对象能被注意)的观察点。

[0165] 这种 3D 视觉注意建模可以与上面关于场景优化所讨论的系统和方法结合。然而,就 3D 视觉注意建模而言,可被限定的视觉目标可以是不同的。例如,视觉目标可被限定成使得 90% 的潜在观察者在穿越酒店大堂时事实上确实是在某一位置观察到特定的数字标牌。

[0166] 以应用与场景优化结合的 3D 视觉注意建模为例,考虑这样的情况:人们可以要对象当从某一观察点观察时保持不显眼,但当从其他观察点观察该对象变得相关时又成为显眼。例如,具有三种不同形式广告内容的酒店,酒店想要其顾客在他们的暂住期间观察并注意所述广告内容。第一内容是为客房升级特色做广告;第二内容是为客房服务特色做广告;第三内容是为酒店舞厅举行表演的门票做广告。这些不同形式的内容在顾客逗留期间的不同时间是相关的。客房升级在顾客登记入住酒店时是相关的,但在其他时候是不相关的。客房服务在顾客进入其房间时是相关的,但当顾客离开酒店时未必是相关的。相比之下,表演门票在几乎任何时候都是相关的。使用与场景优化技术结合的 3D 视觉注意建模技术(上文对二者均有描述),人们可以按下面的方法来优化这些广告材料的布置及内容。首先,可以确定酒店大堂中最可能与信息相关的位置。例如,客房升级在顾客登记入住酒店时相关,这通常在顾客正站在登记柜台面前时发生。靠近电梯的区域可能是一般广告(例如表演门票)的最佳区域。正如酒店可能想要某些标牌从与某些操作任务(例如登记入住、



或进入其房间)对应的观察点显眼,酒店也将想要不相关的标牌不显眼(例如,等待电梯时的客房升级标牌)。可以基于人们站在登记柜台面前时可能处于的位置和取向的集合来分析广告材料的位置和内容。

[0167] 为了测试上述 3D 视觉注意建模和场景优化方法的一个实施例,我们将标准会议室作为我们的测试 3D 场景,其示意图示于图 16。会议室包括桌子 1345、黄色标牌 1350、绿色篮子 1360、电话 1355、紫色标牌 1365 以及人们料想会在会议室中见到的其他典型物品(椅子、废纸篓、屏幕)。人工确定了预计观察者可能会观察房间的代表性地点,产生八个代表性观察位置(观察位置 1366、1377、1378、1379、1380、1381、1382 和 1383)。为此测试,在房间的全部无阻碍(没有家具)的区域中各观察位置以大约 4 英尺的间距隔开。用数字照相机从八个观察位置的每一个拍摄四幅图像,产生 32 幅图像。从观察位置伸出的箭头指示数字照相机为各幅照片瞄准的大致方向,在各观察位置拍摄的各幅照片具有约 90 度的取向差值。记录下 32 幅不同图像各自的位置和取向。

[0168] 然后辨认并标记在 32 幅图像的至少一幅中找到的与 12 个不同对象关联的像素。这通过让用户选择限定包围 32 幅图像每一幅上感兴趣对象的 2D 区域的多边形区域来完成。图 17a 和图 17B 示出 32 幅图像中两幅的艺术效果图,其中多边形包围感兴趣的对象,例如绿色篮子 1360 和黄色标牌 1350。“标记”仅指命名包含对象(例如“黄色标牌”)的区域。应当注意,包围对象的多边形代表标记软件的用途;下面的图像实际上不用多边形加以修饰;一旦由用户指定,所辨认的区域在原始图像上就不以任何方式被符号化。

[0169] 在标记和标签图像后,将图像提交到视觉注意模型以收集图像中模型预测视觉注意将被分配的位置。在提交完所有图像后,计算机为各幅图像记录模型预测视觉注意将被分配的  $x, y$  坐标。计算机还为这些注意注视的每一个计算其是否落入被用户标记和标签的图像区域。计算机还记录包括未被预测受到注意的标记区域(“遗漏”)的各幅图像。所有数据被存储在数据库中,其随后用来生成关于场景内对象明显性的一系列概要。

[0170] 图 18 示出了图表 1395,其说明了为在会议室内标记和标签的 12 个对象 1400 作出的三种不同实例分析。第一分析是对象可见即可观察的可能性( $p(\text{可见})$ )。这是有对象在其中的图像除以图像总数的比率。然后, $p(\text{可见})$  是给出关于感兴趣对象如何被很好地放置在环境内的某种指示的度量。所确定的度量是假定对象可见的情况下其被注意的可能性( $p(\text{注意} | \text{可见})$ ),其通过取得其中对象可见的所有图像并辨认视觉注意模型是否预测到注意会在限定特定对象的区域中发生而计算。所算得的度量是特定对象将被完全注意的可能性( $p(\text{注意})$ ),其通过取得其中模型预测注意分配到对象至少一次的图像数量、然而将该值除以图像的总数而计算。

[0171] 多视角场景分析表示目标对象可以从多个不同距离观察这一事实。例如,考虑上面所述的广告牌实例。人们沿着长而平坦的道路行驶时,有多次机会注意到并因而审视广告牌。相比之下,对于另一广告牌而言,在最后一刻之前可能被山丘或树林遮蔽。通过将上述不同观察点考虑进去,人们可以更准确地分析对象将从其可被观察的不同视角受到视觉注意的可能性。考虑到对象可能受到注意的视角数量,如果没有多视角且仅使用单幅视图,则可能会错误地预测对象会受到或不会受到视觉注意。

[0172] 图 18 示出了可以在由上述分析产生的数据上实现的一小部分可能的评估。从这些数据可以得出多种结论。首先,应当注意紫色标牌对象的可见(图上的白条)频度低于

屏幕对象（也就是说，其在环境内拍摄的图像中较少）。然而，如果我们研究这两个对象的黑条，我们发现当紫色标牌可见时（也就是说，其存在于图像中），其总是被注意（ $p(\text{注意} | \text{可见}) = 1.0$ ），但当屏幕可见时，其从不被注意。这表明尽管紫色标牌位于其将不被经常看见的地方，但模型预测当其出现在视野中时会被分配到注意。

[0173] 紫色标牌对象因此生成前文在酒店大堂的语境中讨论的所期望的特性类型。也就是说，标牌从多个位置不可见（其为不显眼的），但当人们处于其可见的位置（登记柜台旁）时，人们几乎总是注意到该对象。紫色标牌对象可见的概率（白条）约为 15% 的这一事实说明了这一点。但紫色标牌当其可见时将捕获注意的概率（黑条）为 1.0。

[0174] 图 19 是示出包括多视角场景分析系统 1415 的高级功能模块的框图，在各种实施例中，多视角场景分析系统可以完成参照图 15 所述的多视角视觉注意建模的过程。在一个实施例中，高级功能模块存在于计算机系统 408 中，计算机系统包括多个其他功能模块（如操作系统）和硬件，例如存储器或处理器（这两者均未示于图 19）。尽管作为单片机示于图 19，但实践中功能的各部分可以散布在网络构架中的若干或多个计算机中。多视角场景分析系统 1415 包括（前面讨论过的）视觉注意模块 403。其还包括多视角场景分析模块 1425，该模块在需要时调用视觉注意模块 403 来完成感兴趣的分析（前文已参照图 15 的步骤 1340 讨论），并接收分析的结果（包括在哪些图像中的哪些对象被视觉注意模块辨认），并将这些结果或结果的概要存储在数据库 1410 中。数据库 1410 为任何数据存储设备或系统，例如计算机平面文件、计算机存储器或数据库。多视角场景分析系统 1415 还包括图形用户接口模块 1420，其有助于多幅图像（图 15 中步骤 1330 中获得）的输入，然而在此实施例中，有助于图像内感兴趣对象的辨认和标记（图 15 中步骤 1335）。

[0175] 用户 1405 是有兴趣与多视角场景分析系统进行交互的任何人或其他计算机系统。在一个实施例中，用户 1405 是公司雇用的顾问，他对公司拥有或管理的 3D 场景的配置变化作出分析和建议。

[0176] 观察者很多时候会在相同位置中驻留一段合理的时间。例如，某人可能排队等待从杂货店、家居装修店或酒店结帐离开。在这段期间，此人会进行“视觉搜寻”。视觉搜寻是观察者并非寻找任何特定物品而仅仅环顾四周寻找有趣事物的情况。在此视觉搜寻期间，此人将通过移动其眼球注意到不同信息，并且当其眼球到达其转轴边缘时将移动其头部。通常，他会移动其头部和眼球以使感兴趣的项目处于注视的中心。当前技术并不模拟模拟这种在图像上重新对中的动作。相反，当分析图像时，图像的中心总是保持固定。即便当注视点在图像（视野）的边缘处时，也确实如此。在没有重新对中的情况下，当前技术方法将只能够注视图像的边缘，但将从不注视超出该点。相比之下，人类将注意到其视野的边缘并转动其头部，以便其眼球重新对齐到凝视的中心。这使得视觉系统能够在相同方向进行另一次注视。在单幅图像的情况下，由于在图像边缘之外没有更多的信息，人们无法在相同方向进行注视。

[0177] 本文所述的系统和方法（在一些实施例中）使用场景的多幅视图或一幅全景视图来模拟眼球在视觉搜寻期间的重新对中。这可以按照如下方式来实现：

[0178] 1. 从单个观察点生成多幅图像。通过使用 360 度全景摄影机或绕观察轴（垂直轴或其他轴）旋转的多张单幅图像来生成图像。可以获取其中视图彼此“重叠”的多张图像。视图的取向也被指派给各视图。

[0179] 2. 将初始视图（“开始视图”）提供给视觉注意模型。视图可以由某人开始其视觉搜寻的通常观察取向（例如，在商店队列中，可能是“向前”朝收银员看）来确定。也可以从随机选择的取向开始。在全景视图的情况下，可以使用一“片”对处于“开始”视图的全景视图。在多幅视图/图像的情况下，可使用被对中的最靠近开始位置的图像。

[0180] 3. 用视觉注意模型来分析“开始视图”。初始注视通过模型来预测。该注视的取向可被算出（这可用三角法来完成）。如果使用全景视图，则使一“片”新全景视图在该新注视上对中。如果使用多幅图像，则可使用被对中的最靠近该新注视的图像。

[0181] 4. 在新视图的情况下，系统能够对下一显著区域进行分析。

[0182] a. 该过程随后重复（确定注视，然后对中观察点）。

[0183] 实例

[0184] 下面提供了若干非限制性实例，说明上述系统和方法如何可被投入实际使用。

[0185] 实例 1：使用外部变动的稳健性计算

[0186] 背景：酒店业主想要在酒店大堂安装两个数字标牌。她想要这些标牌被顾客注意到，并可以将这些标牌放在 3 个可能的位置中的任一位置上，产生两个数字标牌的 3 种可能配置（也就是说，标牌在位置 1-2、1-3 或 2-3 处）。计算稳健性以推荐将被注意的最佳标牌位置。

[0187] 1. 用数字照相机拍摄酒店大堂的图像，将图像下载到能够运行注意模型的计算机上。计算机安装有视觉注意建模软件（例如，Koch & Itti）以及 Matlab™（可得自 MathWork(Natick, Massachusetts)）。

[0188] 2. 修改图像以包括模拟的数字标牌。生成三幅修改图像，各自模拟两块数字标牌以产生两块标牌在三个可能的位置中的所有组合。使用诸如 Photoshop™（可得自 Adobe 公司 (San Jose, CA)）之类的标准数字照片处理程序。各个模拟数字标牌正确地按比例制作并具有模拟内容，例如酒店标志的图形。将由与三个数字标牌位置中每一个关联的像素地址限定的图像区域存储在电脑文件中。

[0189] 3. 通过注意模型运行修改图像。其输出将包括修改图像的预测显著区域。各个显著区域与存储在步骤 2 中的数字标牌像素地址相比较。如果显著区域落入存储像素地址内或与其重叠，则所预测的注意来到所需位置。这三幅修改图像的每一幅显示数字标牌处于前十位注视中，从而证实这三个位置中任一位置均为良好候选位置。

[0190] 4. 捕获相同场景的多幅图像，不是使用一系列静止照片，就是使用视频和来自视频流的采样图像。图像在 16 小时期间每 5 分钟拍摄一次，从而捕获由多种照明条件和行人运动引起的场景外部变动。目标是让标牌位置对这些类型的变动（照明和行人运动）具有稳健性。将这些图像载入计算机并用模拟数字标牌对其进行修改，如步骤 2 所述。

[0191] 5. 步骤 4 的各幅修改图像用注意模型进行分析，并与存储的像素地址相比较，如步骤 3 所述。这一系列与标牌位置 1 和 2 关联的修改图像表明在 20% 的图像中所预测的注视来到这两个数字标牌位置。类似地，位置 1 和 3 让 35% 的注视来到这两个标牌位置，而位置 2 和 3 让 85% 的注视来到这两个标牌位置。让人将标牌安装在位置 2 和 3 触以得到最具稳健性的配置，从而为酒店提供最佳方案。向酒店业主推荐此方案。

[0192] 实例 2：使用内部变动的稳健性计算

[0193] 背景：酒店业主想要在酒店大堂安装两个数字标牌。她想要这些标牌被顾客注意

到,并可以将这些标牌放在 3 个可能的位置之一处,产生两个数字标牌的 3 种可能配置(也就是说,标牌在位置 1-2、1-3 或 2-3 处)。计算稳健性以推荐将被注意的最佳标牌位置。

[0194] 1. 使用数字照相机拍摄酒店大堂的图像,将图像下载到能够运行注意模型的通用计算机上。计算机安装有视觉注意建模软件(例如, Koch & Itti)以及 Matlab™(可得自 MathWork(Natick, Massachusetts))。

[0195] 2. 修改图像以包括模拟的数字标牌。生成三幅修改图像,各自模拟两块数字标牌以产生两块标牌在三个可能的位置中的所有组合。使用诸如 Photoshop™(可得自 Adobe 公司(San Jose, CA))之类的标准数字照片处理程序。各个模拟数字标牌正确地按比例制作并具有模拟内容,例如酒店标志的图形。将由与三个数字标牌位置中每一个关联的像素地址限定的图像区域存储在电脑文件中。

[0196] 3. 通过注意模型运行修改图像。其输出将包括修改图像的预测显著区域。各个显著区域与存储在步骤 2 中的数字标牌像素地址相比较。如果显著区域落入存储像素地址内或与其重叠,则所预测的注意来到所需位置。这三幅修改图像的每一幅表明数字标牌处于前十位注视中,从而证实三个位置中的任一位置均为良好候选位置。

[0197] 4. 从 Koch & Itti 基础视觉注意模型开始,如步骤 1 所规定。规定模型变化的数量以用于分析修改图像(例如 100 种模型变化)。各个视觉注意模型具有三种不同的特征映射(颜色、取向和亮度);作为这些映射中每一种的加权组合计算出显著性映射。基础视觉注意模型将各种映射的加权参数设为相等(1, 1, 1)。为了产生 100 种模型变化,为各个模型随机设定加权向量。这通过随机设定各个权重并将权重的总和归一化为  $3(3 * (\text{随机权重} / (\text{随机权重} \text{ 总和}))$  的算法来完成。

[0198] 5. 通过 100 种视觉注意模型变化(由 100 个随机加权值限定)来分析各幅图像,并将结果与步骤 2&3 所述的存储像素地址相比较。这一系列与数字标牌位置 1 和 2 关联的修改图像表明在 20% 的图像中所预测的注视来到这两个数字标牌位置。类似地,位置 1 和 3 让 35% 的注视来到这两个标牌位置,而位置 2 和 3 让 85% 的注视来到这两个标牌位置。让人将标牌安装在位置 2 和 3 可得到酒店的最具稳健性的配置。向酒店业主提供此推荐方案。

### [0199] 实例 3:场景优化

[0200] 背景:酒店业主想要在视觉上优化其大堂以及在两块数字标牌上显示的内容。她的具体的视觉目标是让顾客注意到四个目标对象:第一和第二数字标牌、为酒店餐厅做广告的静态图像标牌以及在登记柜台后的员工。

[0201] 1. 为了生成优化选项的分数,对于将注意吸引到目标对象的变化给予回报,并且将真实世界成本与容许的变化相关联。将有关劳动力的估计美元成本以及供应成本分配到被考虑的潜在变化:

[0202] 移动目前位于登记柜台后面的绘画:\$100,

[0203] 改变餐厅标牌后面的照明:\$2500, 以及

[0204] 重新设计在两块数字标牌上显示的内容:每块 \$250 美元。

[0205] 分配给实现视觉目标的回报值如下:

[0206] 将注意吸引到两块数字标牌:每块 \$500,

[0207] 将注意吸引到餐厅标牌:\$250,

[0208] 以及将注意吸引到登记柜台后面的员工 :\$150。

[0209] 2. 使用数字照相机拍摄现有酒店大堂的图像,将图像下载到能够运行注意模型的计算机上。计算机具有视觉注意建模软件,如 Koch & Itti 以及 Matlab™(可得自 MathWork(Natick, Massachusetts))。

[0210] 3. 修改图像以反映被考虑的变化,以产生与可能变化的所有可能组合相关的多幅图像。使用诸如 Photoshop™(可得自 Adobe 公司 (San Jose, CA)) 之类的标准数字照片处理程序。还指定了与顾客视觉目标相关的目标对象的像素地址,并将其存储在存储器中。

[0211] 4. 使用注意模型分析步骤 3 的各幅图像,并将由模型预测的显著对象与目标对象的存储像素地址相比较。通过从实现修改图像中的视觉目标的回报值中减去变化的成本来计算各个模拟配置的分数的,该视觉目标的实现由预测视觉注意与目标对象的像素地址的重叠表示。例如,当注意被分配到餐厅标牌时利用了移动绘画这一变化,其分数是  $\$250 - \$100 = \$150$ 。在分析所有模拟图像之后,找到的最具成本效益的方案是以 \$100 的成本移动绘画并以 \$250 的成本修改内容之一的颜色(总成本 \$350)。这些变化使所有视觉目标得以实现,产生 \$1400 的回报分数和 \$1050 的总分数。

#### [0212] 实例 4:多视角场景分析

[0213] 背景:从实例 3 继续进行,已作了所推荐的变化。酒店业主想要了解从酒店大堂中多个视角观察的各个目标对象的视觉显著性。

[0214] 1. 辨认分布于整个大堂的四个感兴趣位置,并通过以 90 度增量旋转摄影机从各个位置获取四幅数字照片,产生总共 16 幅图像,各幅图像代表一个视角。图像是用数字照相机拍摄的照片。将图像下载到能够运行注意模型的计算机上。计算机安装有视觉注意建模软件,例如 Koch & Itti 以及 Matlab™(MathWork(Natick, Massachusetts))。对于各个视角,辨认目标对象的像素地址,并将其存储在计算机的存储器中,并且用标识符标记目标对象。

[0215] 2. 使用注意模型分析步骤 1 的 16 幅图像中的每一幅。对于各幅图像,由视觉注意模型确定哪些目标对象是可见的以及哪些目标对象被预测可吸引注意。

[0216] 3. 计算各个目标对象在所有图像上可见的概率,并且还计算其受到注意的概率。这些数据出现于给酒店业主的报告中,提供从多个视角对大堂中视觉特性的更好理解。

[0217] 应当注意,本文描述的方法和系统的应用实例还可在所讨论特定应用之外得到广泛应用。例如,这些应用包括零售环境。

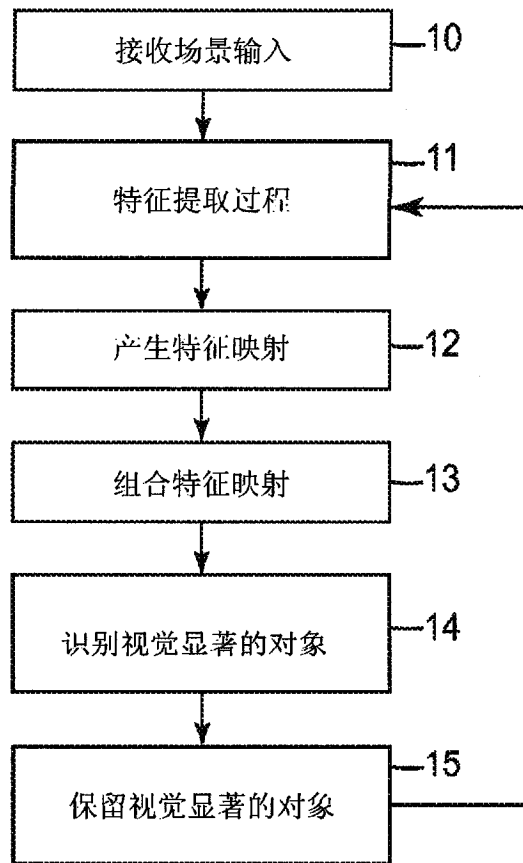


图 1

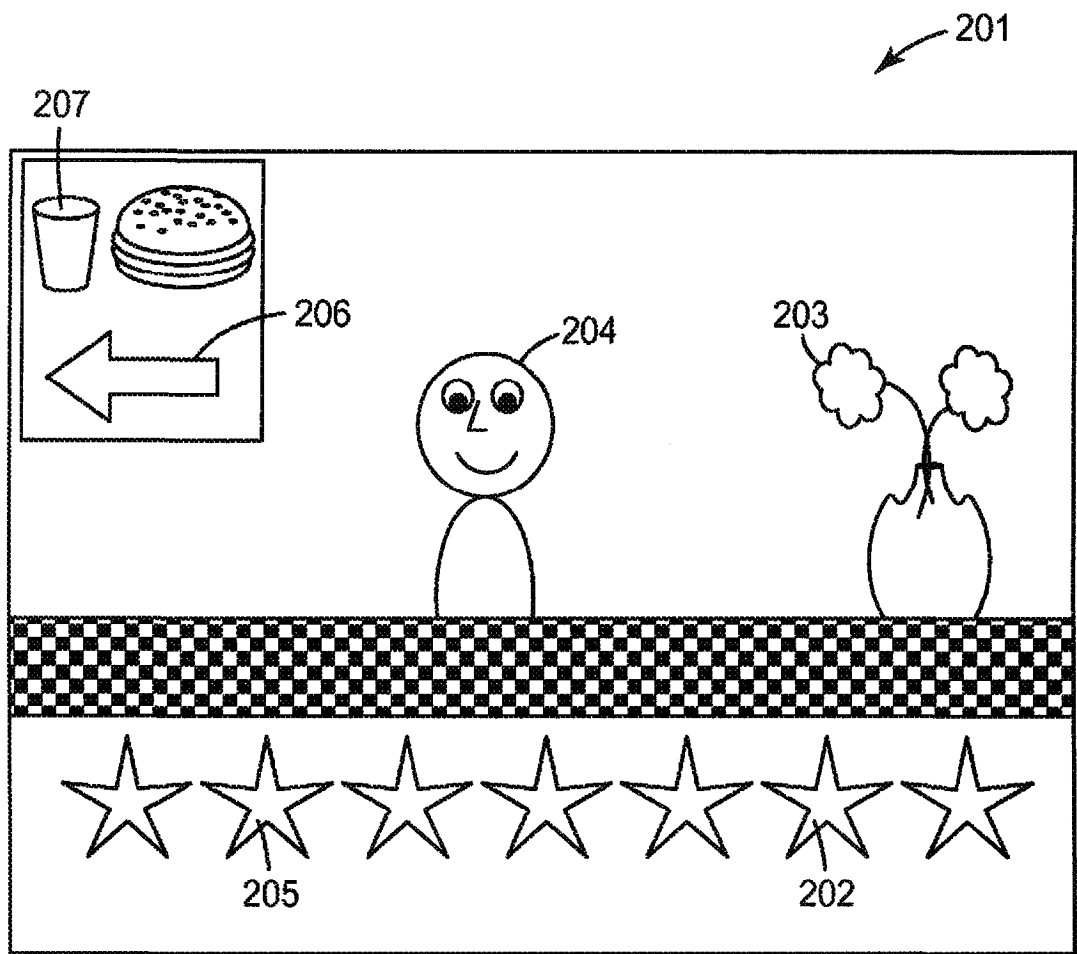


图 2

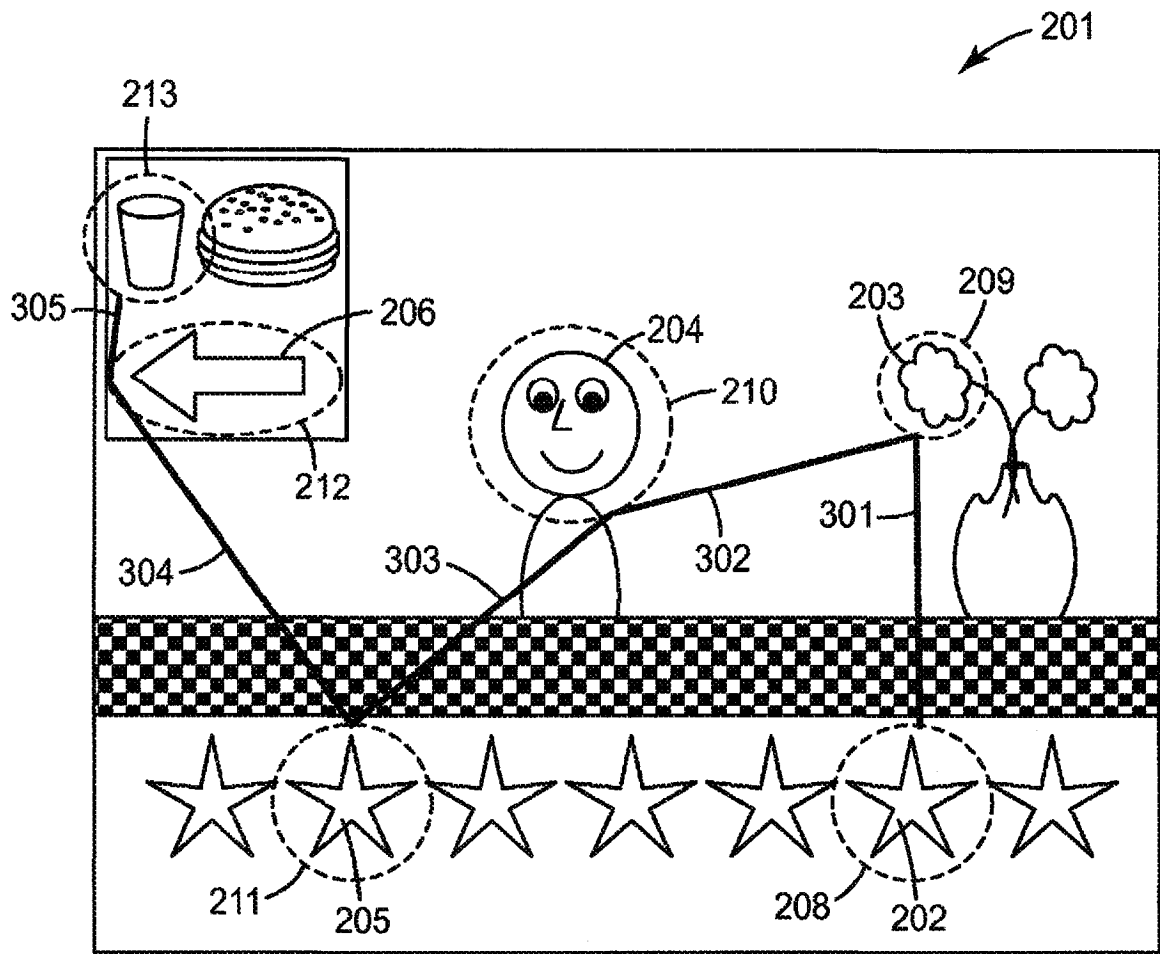


图 3A





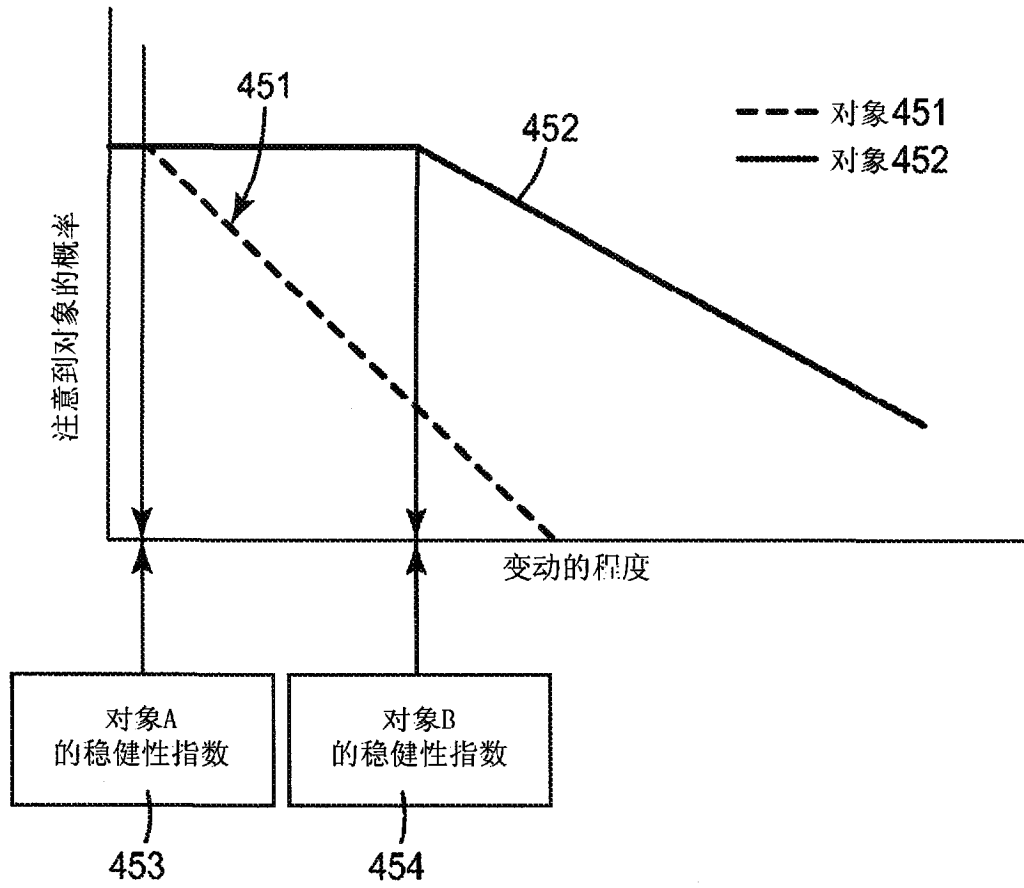


图 4

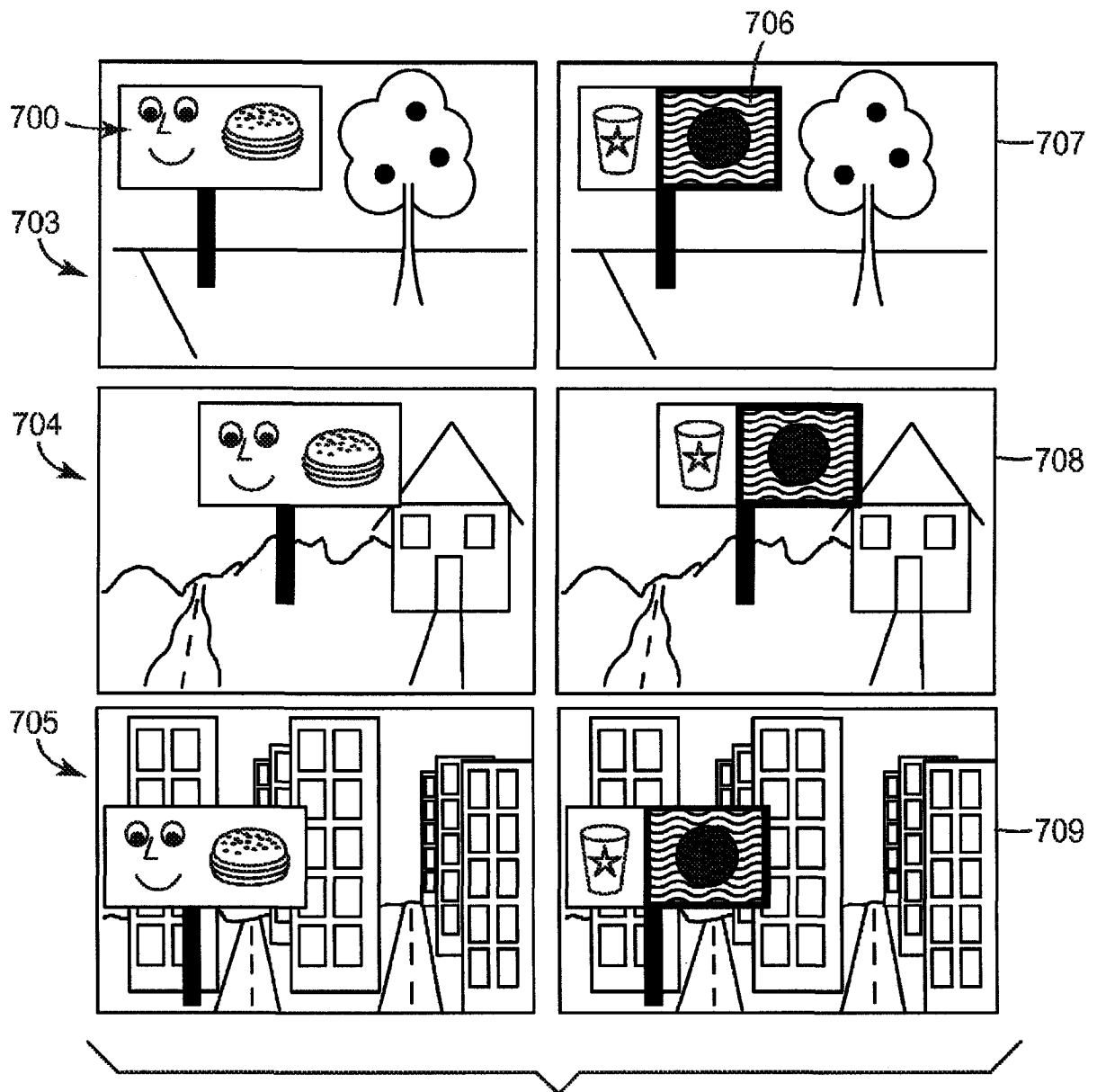


图 5

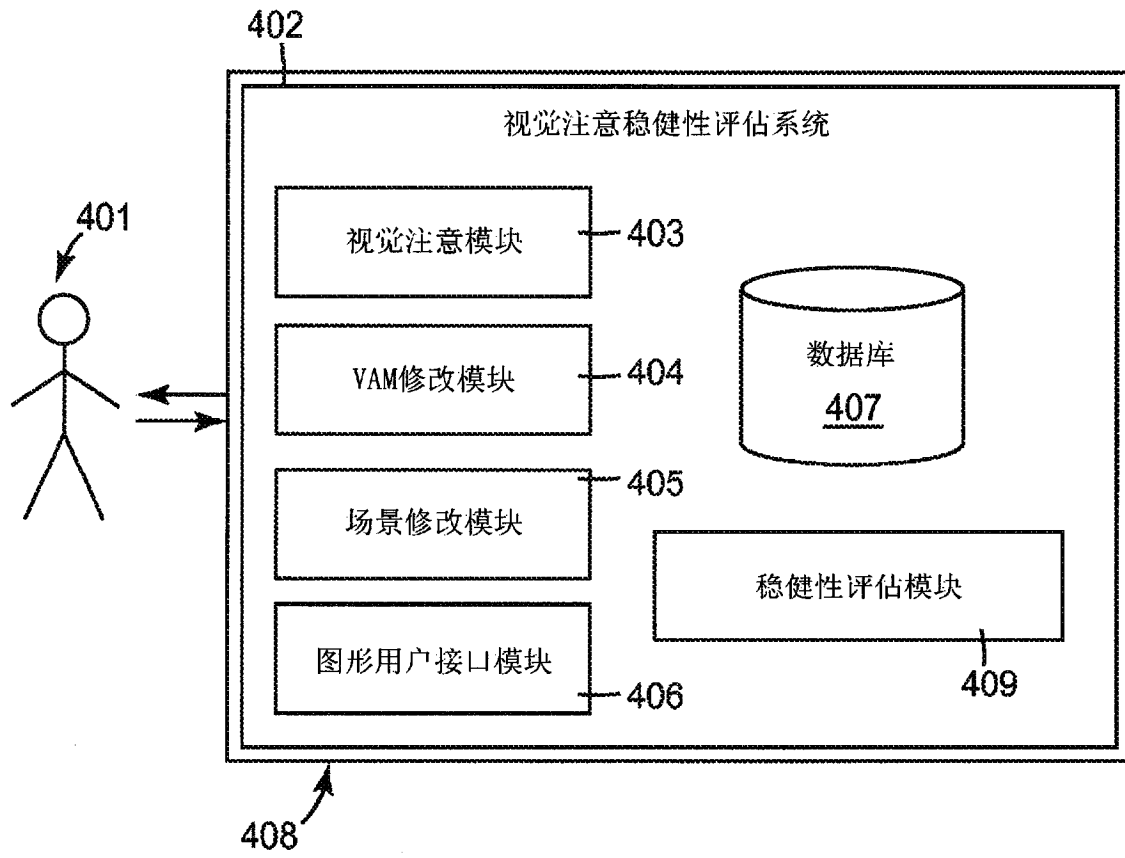


图 6

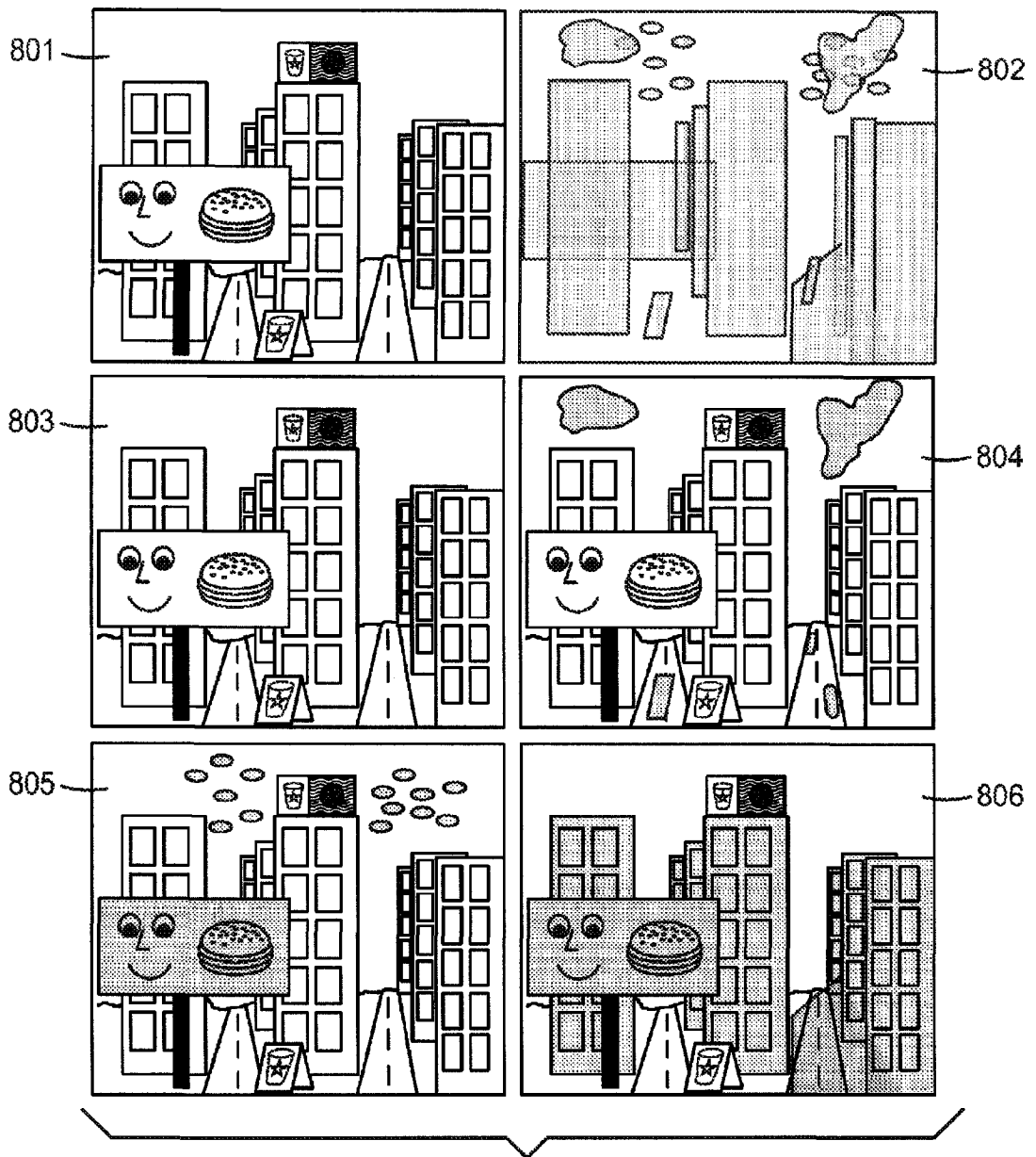


图 7

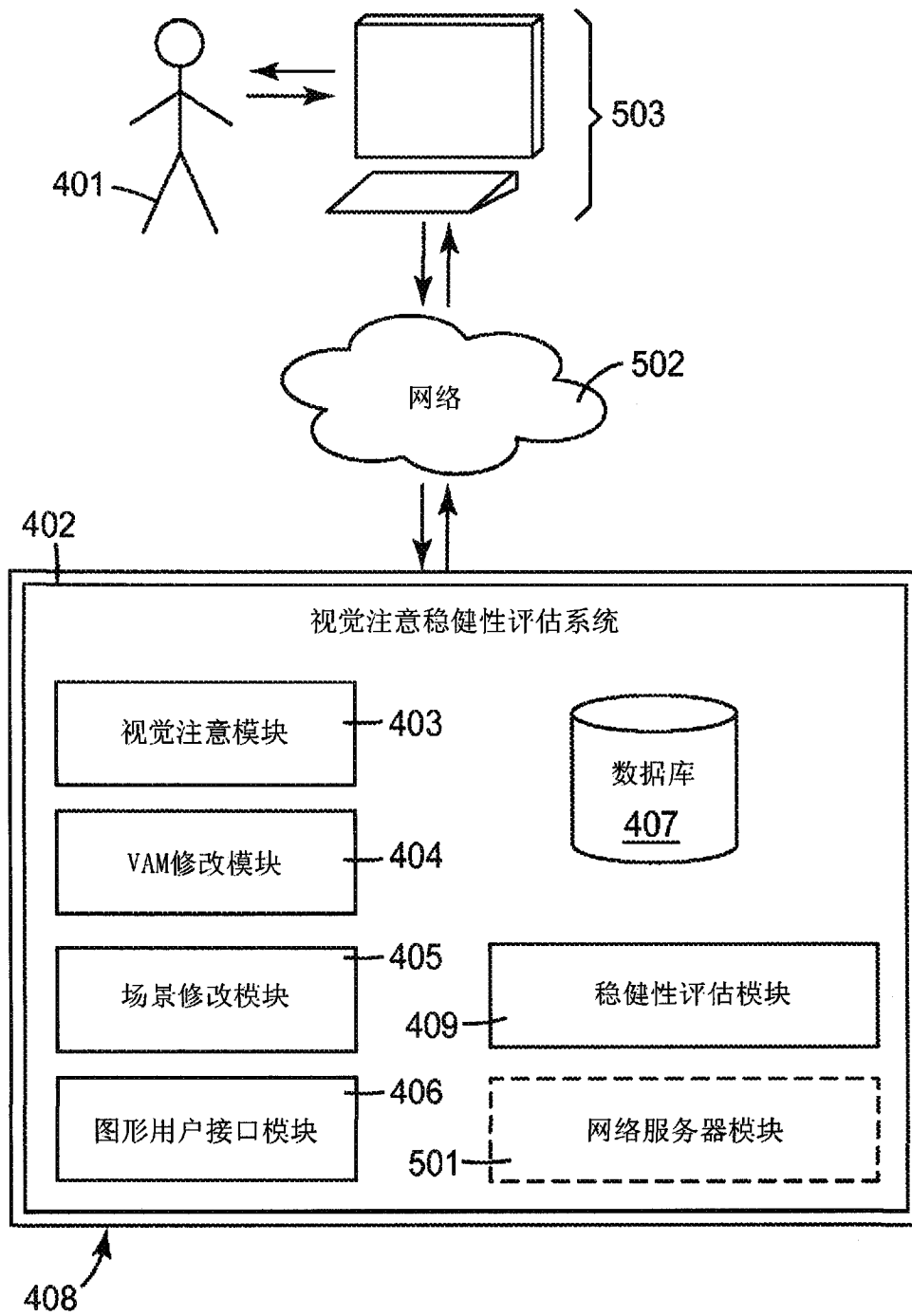


图 8

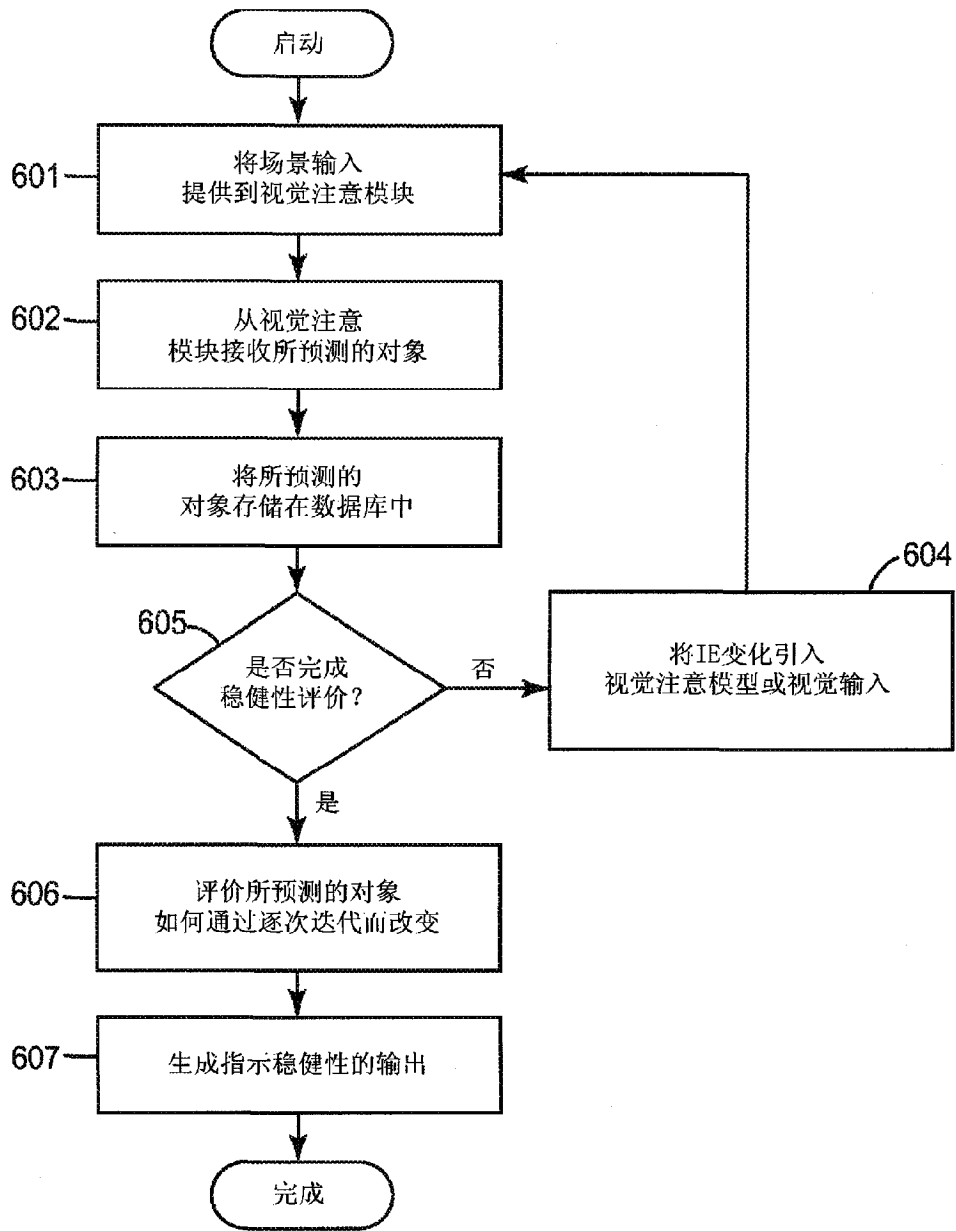


图 9

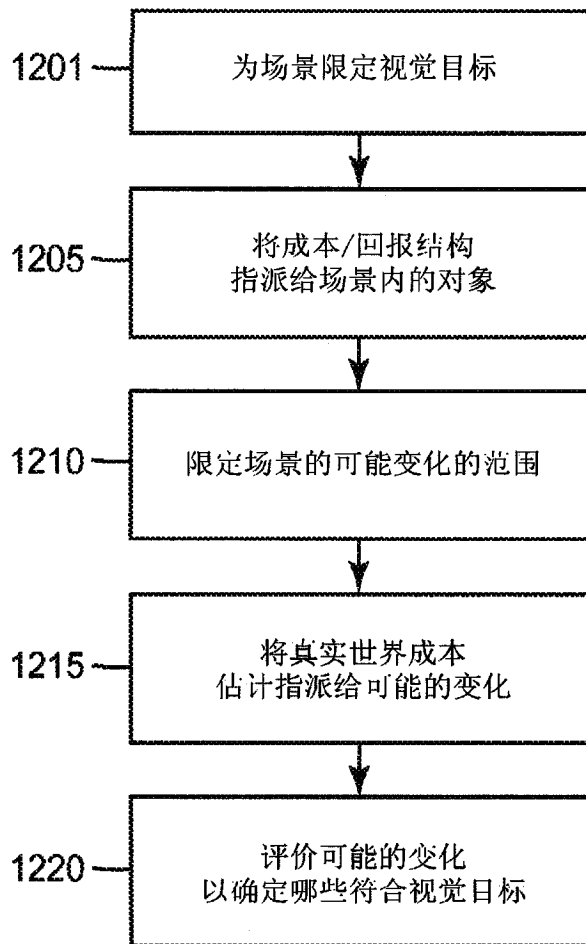


图 10



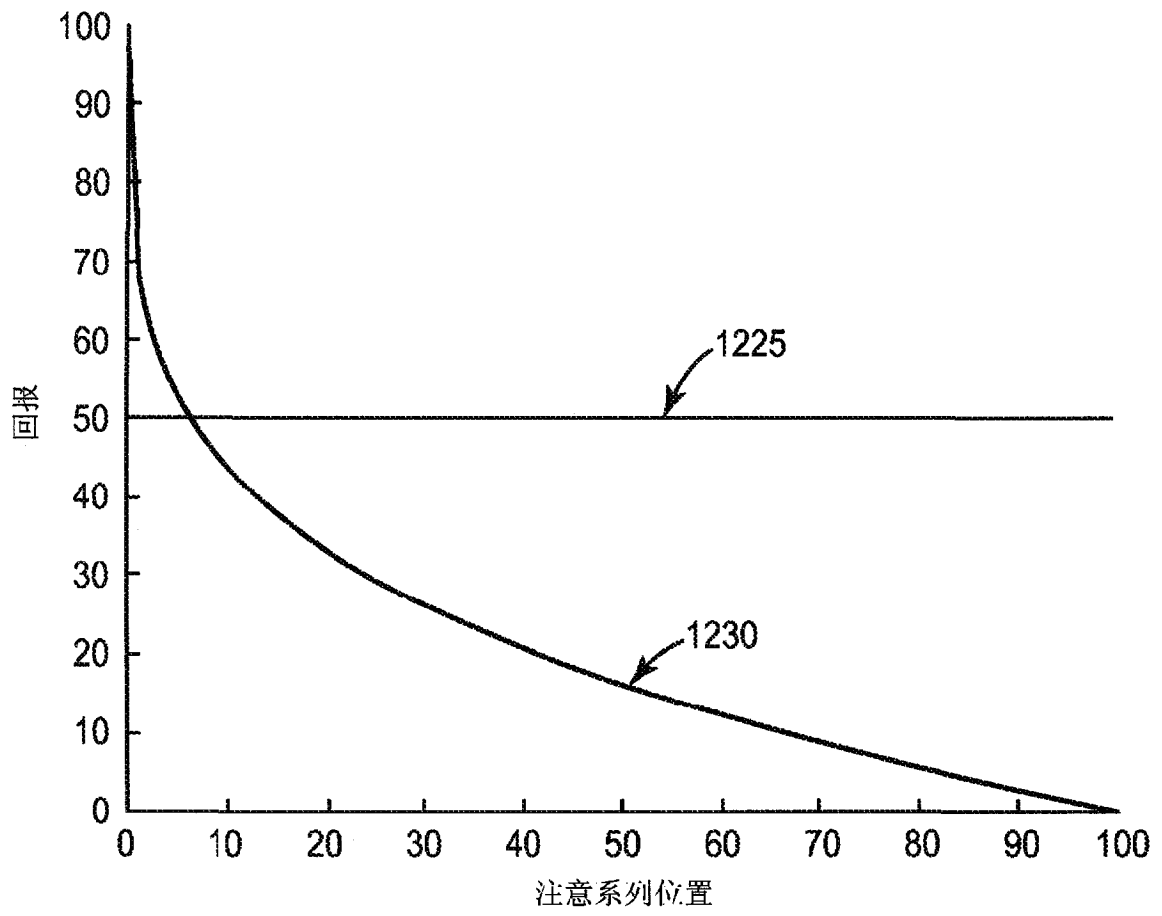


图 11

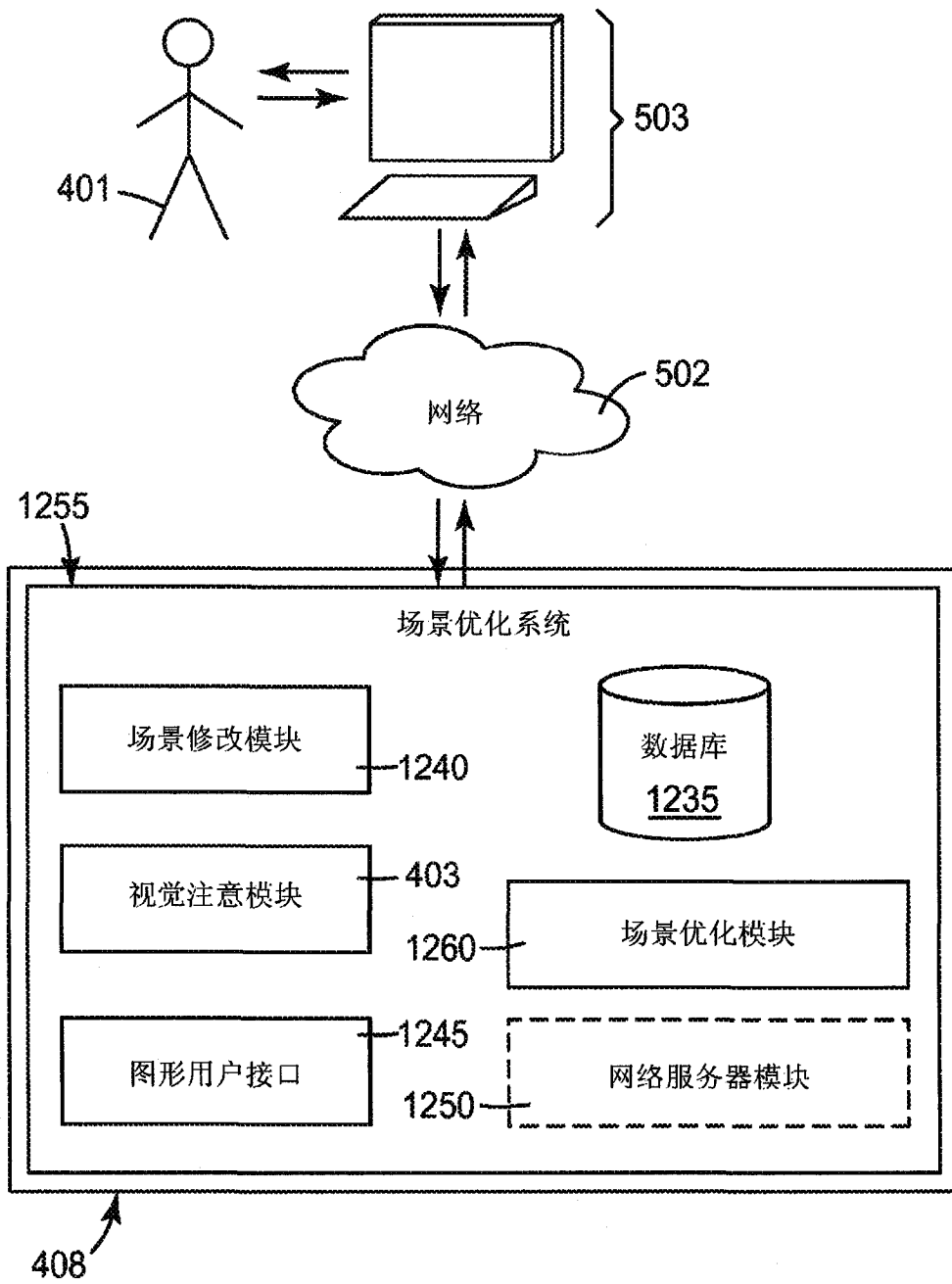


图 12

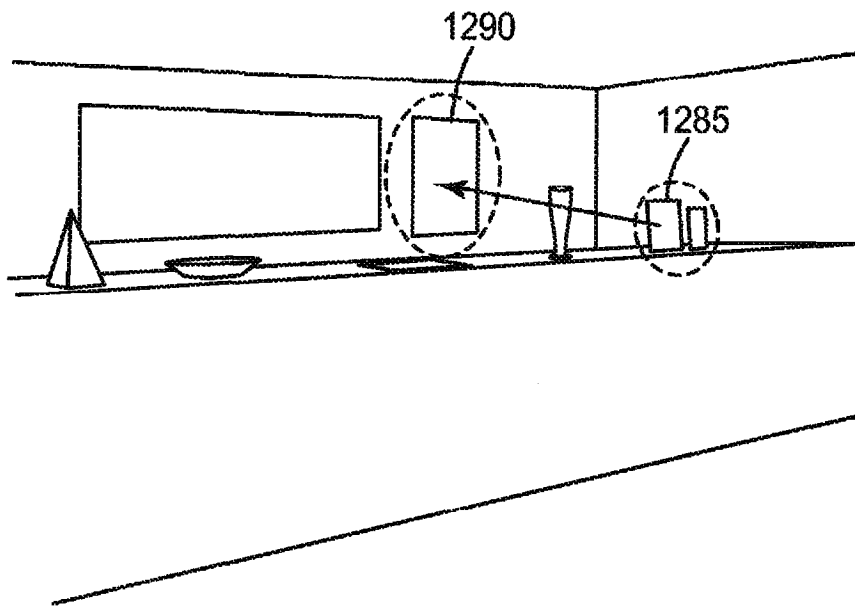


图 13A

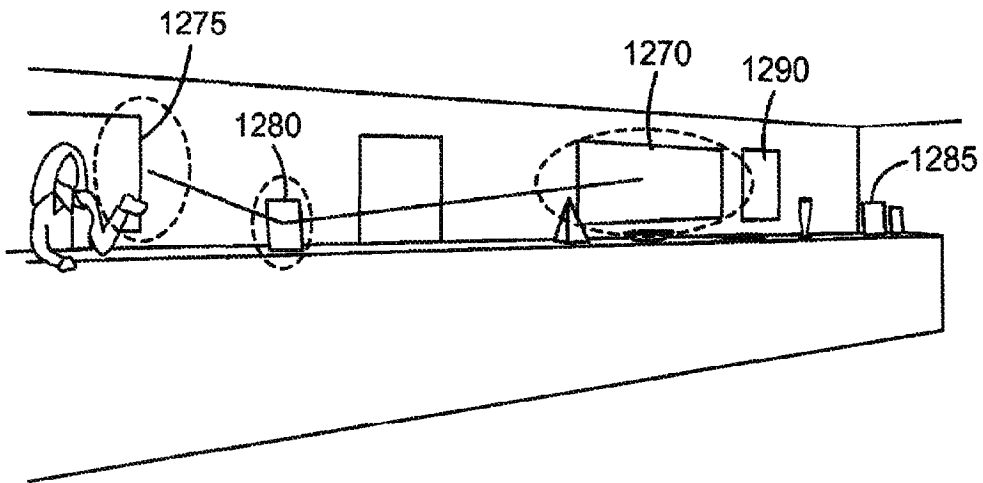


图 13B

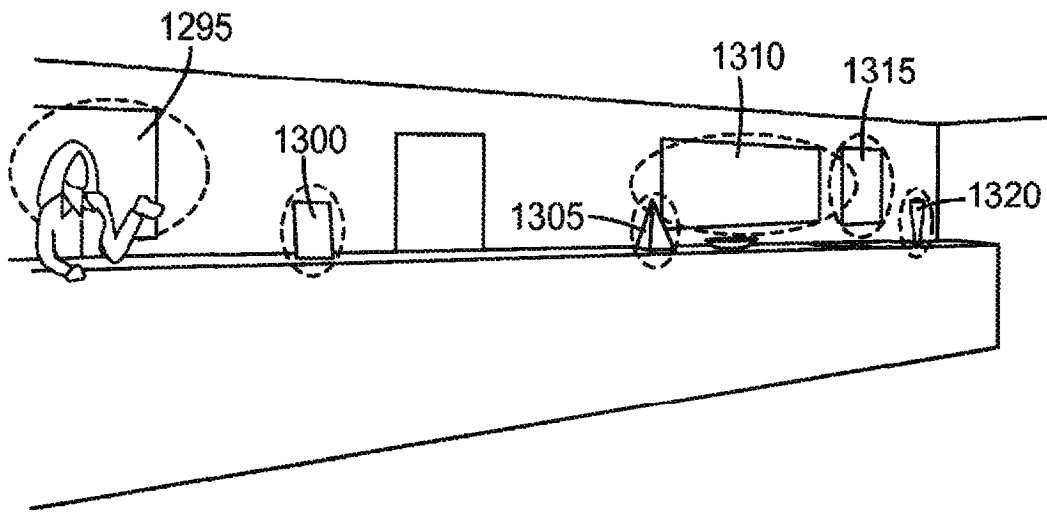


图 14A

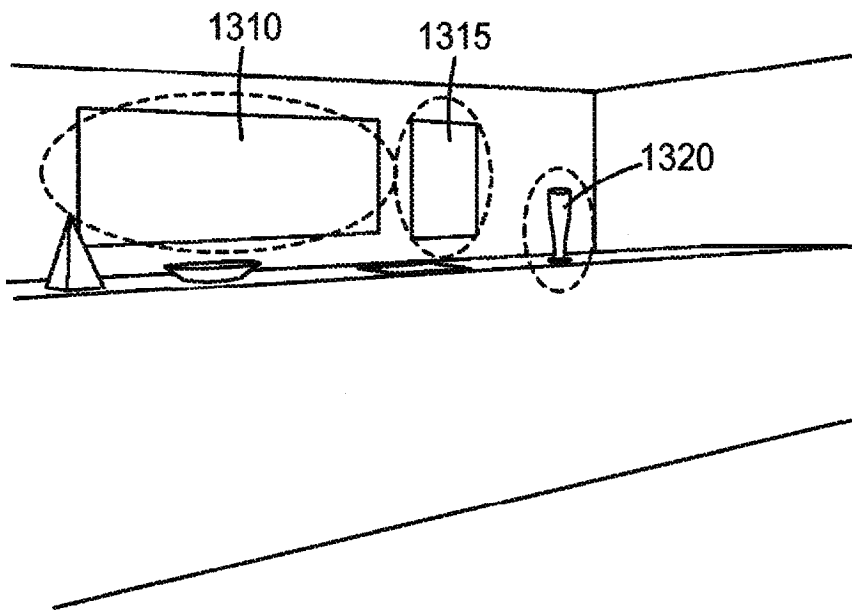


图 14B

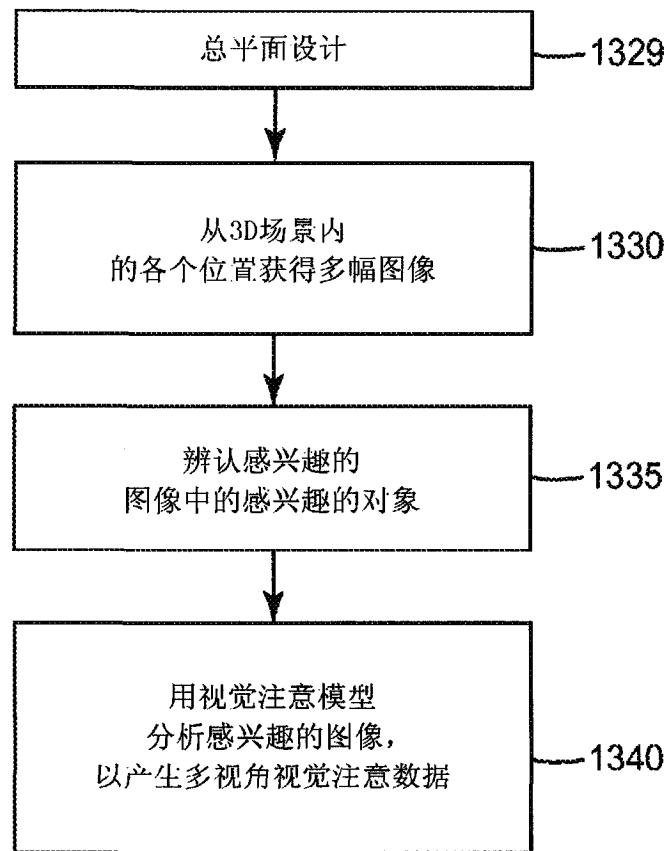


图 15

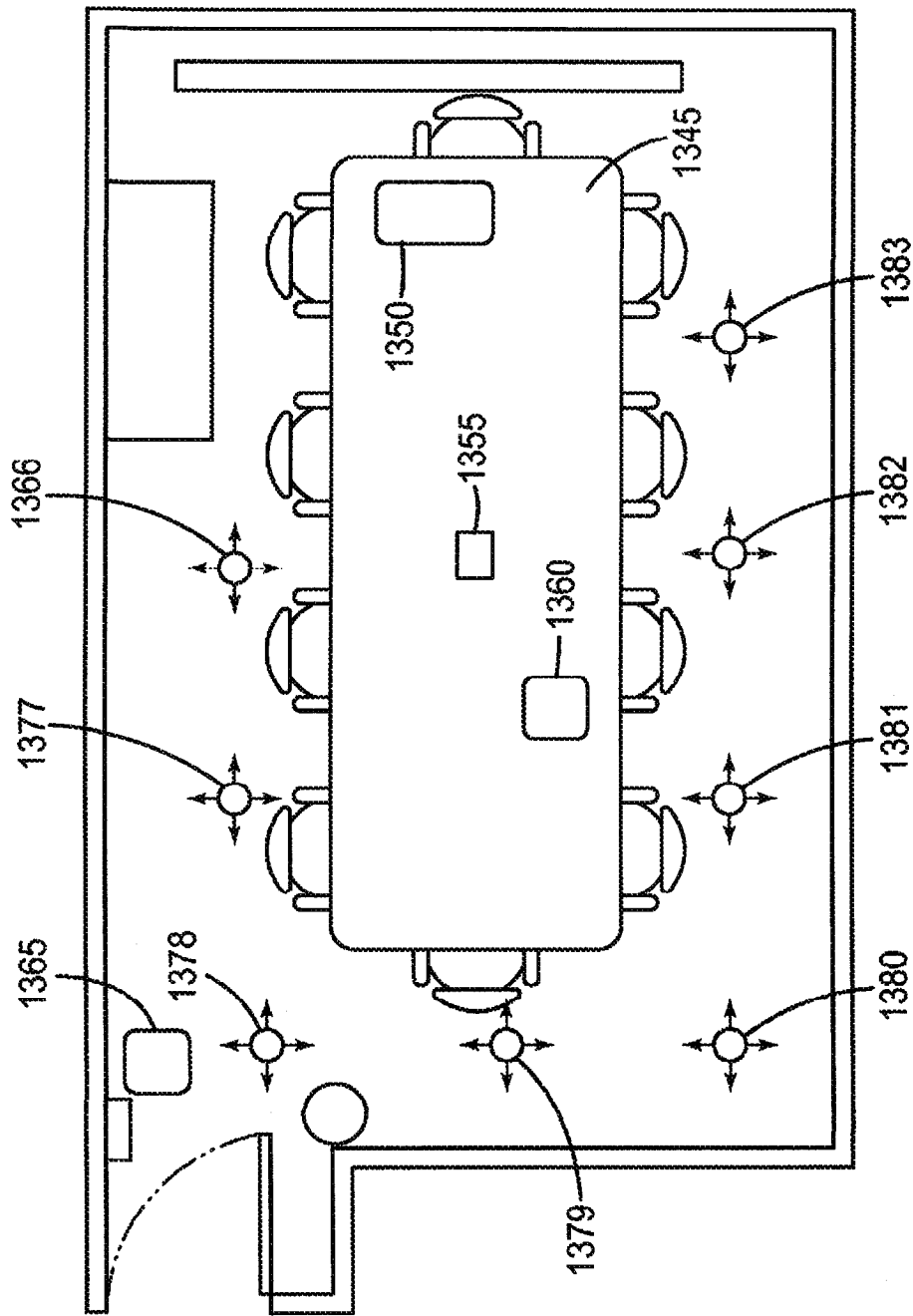


图 16

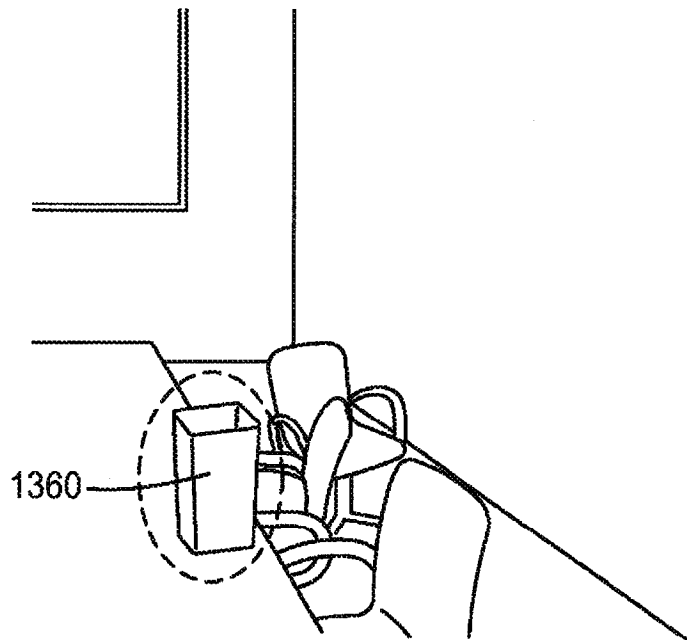


图 17A

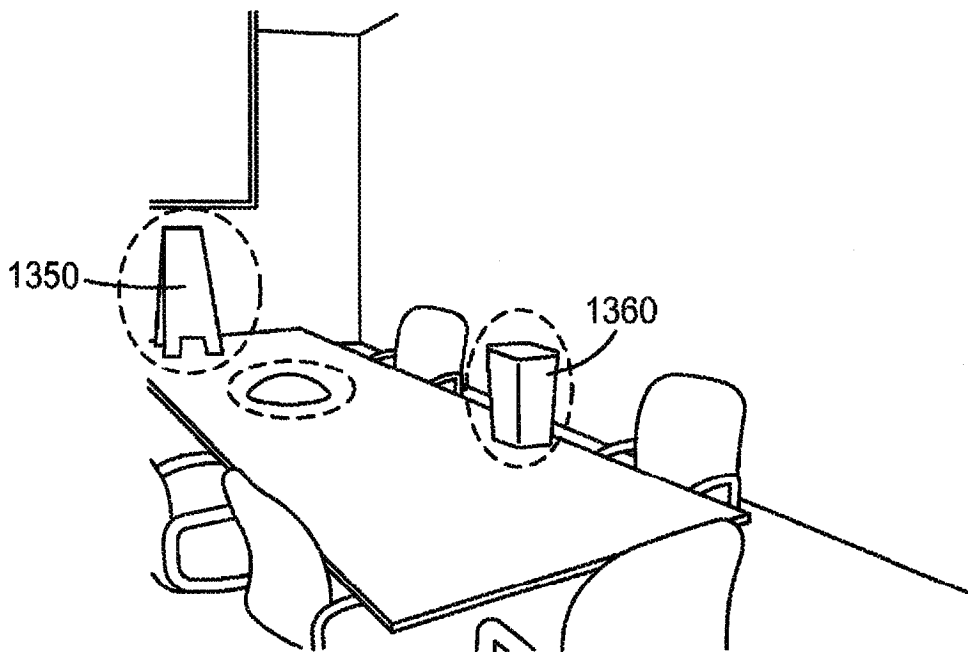


图 17B

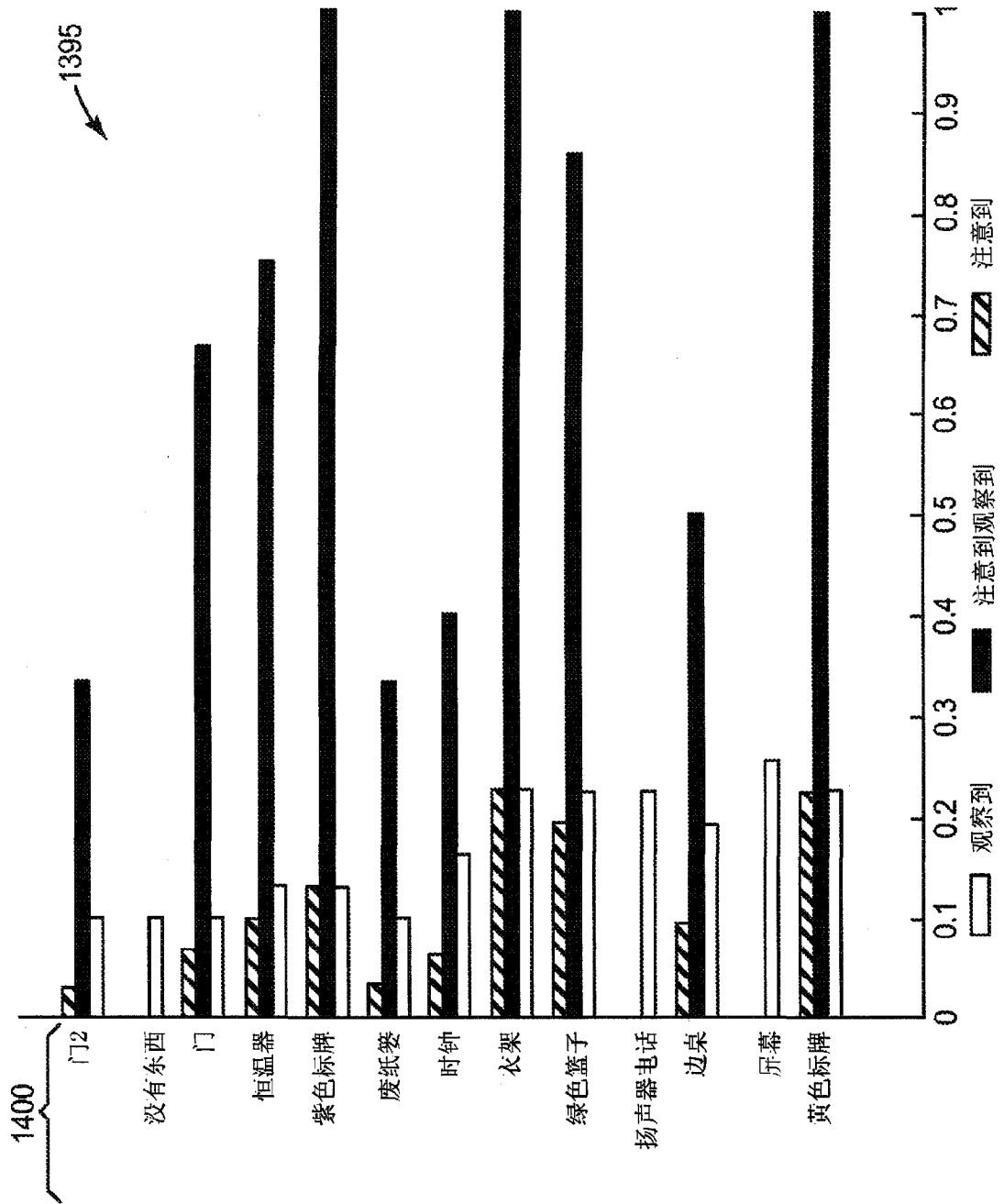


图 18



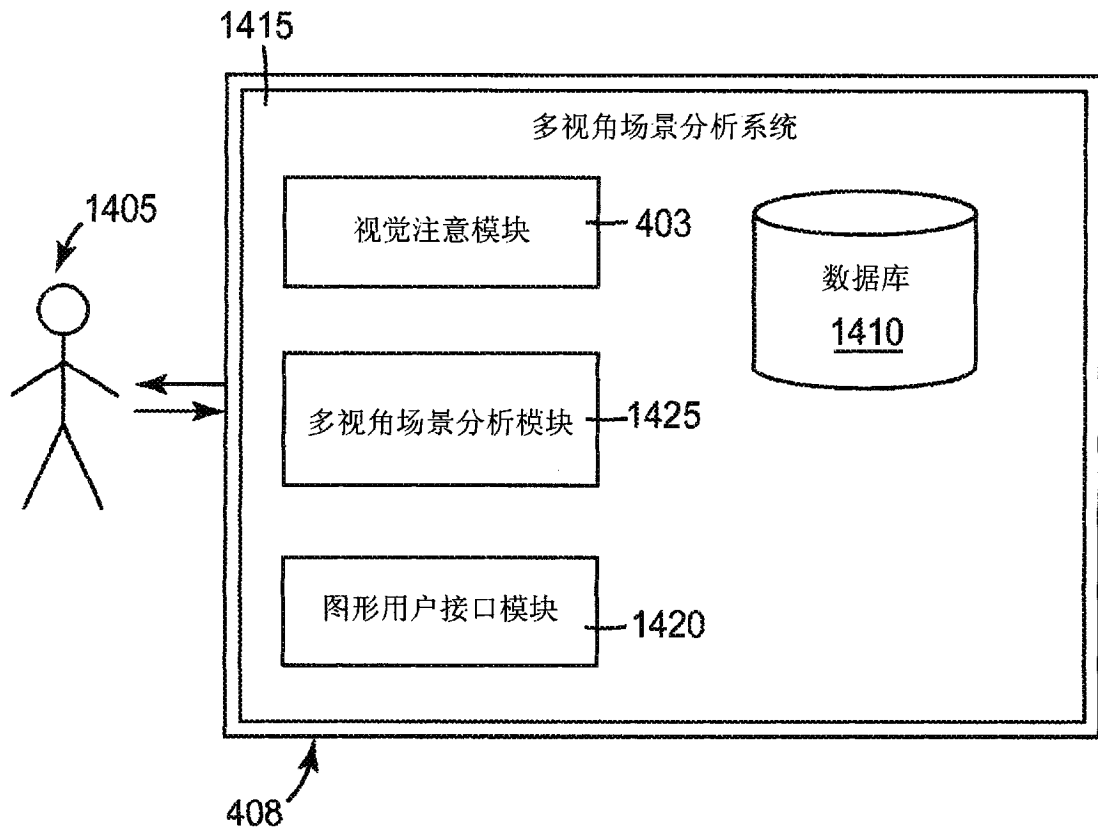


图 19