(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2020 年 6 月 25 日 (25.06.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/125499 A1

(51) 国际专利分类号:
G06F 3/01 (2006.01)

(21) 国际申请号: PCT/CN2019/124352

(22) 国际申请日: 2019 年 12 月 10 日 (10.12.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811543901.X 2018年12月17日 (17.12.2018) CN

(71) 申请人: 中国科学院深圳先进技术研究院 (**SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY CHINESE ACADEMY OF SCIENCES**) [CN/CN]; 中国广东省深圳市南山区深圳大学城学苑大道 1068 号, Guangdong 518000 (CN)。

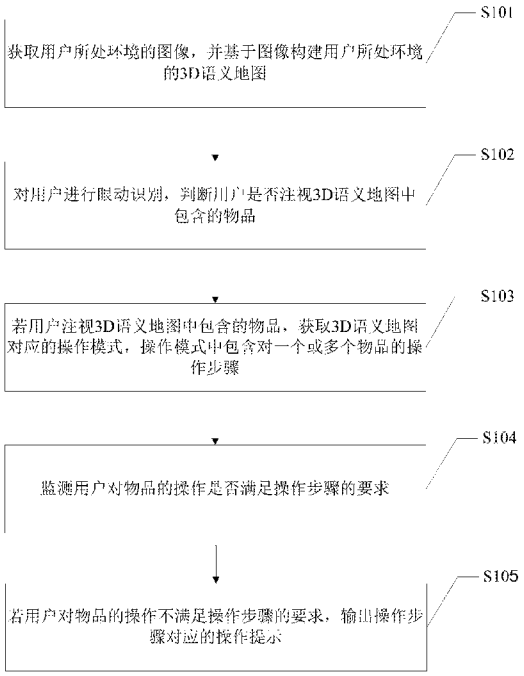
(72) 发明人: 程俊 (**CHENG, Jun**); 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518000 (CN)。 王鹏 (**WANG, Peng**); 中国广东省深圳市南山区深圳大学城学苑大道1068号, Guangdong 518000 (CN)。

(74) 代理人: 深圳中一联合知识产权代理有限公司 (**SHENZHEN ZHONGYI UNION INTELLECTUAL PROPERTY AGENCY CO., LTD.**); 中国广东省深圳市福田区园岭街道深南中路1014号报春大厦9楼（5号信箱）, Guangdong 518028 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,

(54) **Title:** OPERATION PROMPTING METHOD AND GLASSES

(54) 发明名称: 一种操作提示方法及眼镜



S101 ACQUIRE AN IMAGE OF AN ENVIRONMENT WHERE A USER IS LOCATED, AND CONSTRUCT, ON THE BASIS OF THE IMAGE, A 3D SEMANTIC MAP OF THE ENVIRONMENT WHERE THE USER IS LOCATED

S102 IDENTIFY THE EYE MOVEMENT OF THE USER, SO AS TO DETERMINE WHETHER THE USER LOOKS AT AN ARTICLE INCLUDED IN THE 3D SEMANTIC MAP

S103 IF THE USER LOOKS AT AN ARTICLE INCLUDED IN THE 3D SEMANTIC MAP, ACQUIRE AN OPERATION MODE CORRESPONDING TO THE 3D SEMANTIC MAP, THE OPERATION MODE COMPRISING OPERATION STEPS FOR ONE OR MORE ARTICLES

S104 MONITOR WHETHER THE OPERATION ON THE ARTICLE BY THE USER SATISFIES THE REQUIREMENTS OF THE OPERATION STEPS

S105 IF THE OPERATION ON THE ARTICLE BY THE USER DOES NOT SATISFY THE REQUIREMENTS OF THE OPERATION STEPS, OUTPUT AN OPERATION PROMPT CORRESPONDING TO THE OPERATION STEPS

图 1

(57) **Abstract:** An operation prompting method and apparatus, and glasses, being applicable in the technical field of data processing. Said method comprises: acquiring an image of an environment where a user is located, and constructing, on the basis of the image, a 3D semantic map of the environment where the user is located (S101); identifying the eye movement of the user, so as to determine whether the user looks at an article included in the 3D semantic map (S102); if the user looks at an article included in the 3D semantic map, acquiring an operation mode corresponding to the 3D semantic map, the operation mode comprising operation steps for one or more

[见续页]

WO 2020/125499 A1

CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

articles (S103); monitoring whether the operation performed on the article by the user satisfies the requirements of the operation steps (S104); and if the operation performed on the article by the user does not satisfy the requirements of the operation steps, outputting an operation prompt corresponding to the operation steps. A user can learn a problem in his/her own operation in a timely manner, without the need to perform any manual input or actively watch a screen all the time, realizing intelligent matching of an operation tutorial of the user, greatly improving the efficiency of determining an operation error of the user.

(57) 摘要：一种操作提示方法、装置及眼镜，适用于数据处理技术领域，该方法包括：获取用户所处环境的图像，并基于图像构建用户所处环境的3D语义地图（S101）；对用户进行眼动识别，判断用户是否注视3D语义地图中包含的物品（S102）；若用户注视3D语义地图中包含的物品，获取3D语义地图对应的操作模式，操作模式中包含对一个或多个物品的操作步骤（S103）；监测用户对物品的操作是否满足操作步骤的要求（S104）；若用户对物品的操作不满足操作步骤的要求，输出操作步骤对应的操作提示。用户无需再进行任何手动操作输入，也无需一直主动对着屏幕观看，即可以及时得知自己操作中的问题，实现了对用户操作教程的智能匹配，极大地提升了用户对操作错误的确定效率。

一种操作提示方法及眼镜

技术领域

本发明属于数据处理技术领域，尤其涉及操作提示方法及眼镜。

背景技术

实际生活和工作中，用户经常需要上网搜索一些操作教程来指导自己的一些操作是否正确，例如搜索一些做饭图文教程并根据做饭图文教程来看看自己烹饪方法是否有误，或者搜索一些设备操作教程来判断设备操作是否有误，现有技术中，用户都是通过电脑手机等设备手动输入一些关键词来进行搜索，得到对应的图文视频教程，再根据这些图文视频教程来逐步进行对比，但这些都需要用户进行大量的手动操作输入，且需要用户一直主动对着屏幕进行对比，确定操作的错误，操作繁琐且效率低下。

技术问题

有鉴于此，本发明实施例提供了一种操作提示方法及眼镜，以解决现有技术中用户需要大量手动操作查看和对比，才能确定自己操作中存在的错误，操作繁琐效率低下的问题。

技术解决方案

本发明实施例的第一方面提供了一种操作提示方法，包括：

获取用户所处环境的图像，并基于所述图像构建用户所处环境的 3D 语义地图；

对用户进行眼动识别，判断所述用户是否注视所述 3D 语义地图中包含的物品；

若所述用户注视所述 3D 语义地图中包含的物品，获取所述 3D 语义地图对应的操作模式，所述操作模式中包含对一个或多个所述物品的操作步骤；

监测所述用户对所述物品的操作是否满足所述操作步骤的要求；

若所述用户对所述物品的操作不满足所述操作步骤的要求，输出所述操作步骤对应的操作提示。

本发明实施例的第二方面提供了一种眼镜，所述眼镜包括存储器、处理器，所述存储器上存储有可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现如上所述的操作提示方法的步骤。

有益效果

本发明实施例与现有技术相比存在的有益效果是：通过构建用户所处环境的 3D 语义地图，并在识别出用户存在物品注视行为，即确定用户存在操作提示需求时，根据 3D 语义地图实际情况智能识别出用户所需的操作模式（即对物品的操作教程），并根据用户对物品的实际操作情况，来进行监测和提示用户操作，使得用户无需再进行任何手动操作输入，也无需一直主动对着屏幕观看，即可以及时得知自己操作中的问题，实现了对用户操

作教程的智能匹配，极大地提升了用户对操作错误的获知效率。

附图说明

为了更清楚地说明本发明实施例中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 是本发明实施例一提供的操作提示方法的实现流程示意图；

图 2 是本发明实施例二提供的操作提示方法的实现流程示意图；

图 3A 和图 3B 是本发明实施例三提供的操作提示方法的实现流程示意图；

图 4A 和图 4B 是本发明实施例四提供的操作提示方法的实现流程示意图；

图 5 是本发明实施例五提供的操作提示方法的实现流程示意图；

图 6 是本发明实施例六提供的操作提示方法的实现流程示意图；

图 7 是本发明实施例七提供的操作提示装置的结构示意图；

图 8 是本发明实施例八提供的眼镜的示意图。

本发明的实施方式

以下描述中，为了说明而不是为了限定，提出了诸如特定系统结构、技术之类的具体细节，以便透彻理解本发明实施例。然而，本领域的技术人员应当清楚，在没有这些具体细节的其它实施例中也可以实现本发明。在其它情况中，省略对众所周知的系统、装置、电路以及方法的详细说明，以免不必要的细节妨碍本发明的描述。

为了说明本发明所述的技术方案，下面通过具体实施例来进行说明。

为了便于理解本发明，此处先对本发明实施例进行简要说明，为了便于用户即使发现自己的操作问题，本发明实施例首先会基于用户环境的图像来构建对应的 3D 语义地图，从而确定出用户实际所处环境的情况以及环境内包含的物体等，但考虑到实际情况中即使用户身处某一环境中也不一定就有操作需求，例如用户可能只是经过某一地点，或者用户虽然身处某一地点之中但并不需要进行操作，因此为了防止误触发，本发明实施例中还会对用户进行眼动分析，判断用户是否存在对物品的注视行为，并在用户存在注视行为时，判定用户存在操作提示的需求，此时在根据用户实际所处的 3D 语义地图情况来智能匹配用户所需的操作教程，最后对用户对物品的实时操作进行分析，并在用户操作不满足操作教程的要求时，对应进行正确的操作步骤提示，从而使得用户无需进行繁琐的手动操作，也可以及时获得所需的操作教程，并及时得知自己操作中存在的问题及正确的操作方式。

应当明确地，在本发明实施例中的具体执行主体可以根据实际应用情况需求的不同而

不同，例如，其既可以是可穿戴设备等智能设备（如智能眼镜），也可以是服务器等设备，当执行主体为可穿戴设备时，本发明实施例中所有的数据采集、处理及输出操作均由可穿戴设备完成，而当执行主体为服务器等无法直接对用户进行数据采集输出的设备时，本发明实施例中对用户数据采集和输出的工作均由其他设备完成，即本发明实施例中数据获取和输出的直接对象不是用户，而是可以对用户进行数据采集输出的其他设备，以实现对用户的操作提示目的，例如，由智能眼镜负责对用户的数据采集并发送给服务器进行处理，由服务器将处理后生成的提示发送至智能眼镜，再由智能眼镜最终输出至用户，具体执行主体的确定，可由研发人员根据实际研发情况和需求来进行选定和设计，此处以执行主体为智能眼镜为例来进行说明，详述如下：

图 1 示出了本发明实施例一提供的操作提示方法的实现流程图，详述如下：

S101，获取用户所处环境的图像，并基于图像构建用户所处环境的 3D 语义地图。

为了实现对用户需求的智能准确识别，本发明实施例中首先会对用户所处环境进行图像采集，并进行 3D 语义地图的构建。其中，3D 语义地图就是包含语义信息的环境三维地图。本发明实施例中，可以在智能眼镜中设置广角摄像头来进行环境图像的采集，在获取到环境图像之后，再基于这些图像进行环境三维地图的构建，并识别其中包含的物品以及每种物品相关属性数据。具体而言，3D 语义地图构建的方法包括但不限于如基于立体视觉的 3D 语义地图算法或者其他构建算法，此处不予限定，具体可由技术人员根据实际需求选取或设定，或者参考本发明的其他相关实施例。

S102，对用户进行眼动识别，判断用户是否注视 3D 语义地图中包含的物品。

当用户身处某一环境内时，并不能说明其就存在操作教程指导的需求，例如用户可能只是经过某一地点，或者只是处于某一地点但并不需要进行操作等，因此，若直接根据用户所处环境来进行操作教程匹配及提示等操作，可能会导致对用户的误提示，但当用户身处某一环境，且持续性地注视一些物品时，说明用户极有可能是需要进行物品操作的，因此，本发明实施例会对用户是否存在对物品的注视行为进行识别，并会在确定出用户存在注视行为时，才进行后续的操作教程匹配等操作，具体而言，本发明实施例中智能眼镜内会内置对用户进行眼动拍摄的装置，例如眼动仪等装置，并由这些装置来对用户进行眼睛图像数据的采集并对采集的眼睛图像进行眼动分析，以确定出用户的眼动情况如何，如用户是否眨眼、注视区域在哪以及是否有注视物品等，在本发明实施例中，具体的眼动分析追踪方法可由技术人员自行选定，既可以使用一些现有的算法，也可根据需求自行设置，或者参考本发明其他相关实施例方法。

在本发明实施例中，需要确认用户是否存在注视物品的行为，即在进行眼动识别时不

仅需要识别出用户看的物品，还需要进一步地确认是否存在持续性的看某个物品（即注视物品）的行为，因此本发明实施例中，在确定出用户看的物品的基础上，还需要统计用户看物品的时间，并依此判断是否注视物品，具体可参考本发明实施例三。

S103，若用户注视 3D 语义地图中包含的物品，获取 3D 语义地图对应的操作模式，操作模式中包含对一个或多个物品的操作步骤。

当用户身处某一环境且持续性地注视一些物品时，说明用户极有可能是需要进行物品操作的，此时，本发明实施例会对用户环境的 3D 语义地图进行分析确定其实际所处环境情况，如是厨房还是设备操作室等，再根据环境情况来估计用户可能进行的操作以及对应所需的操作教程。本发明实施例中，操作教程既可以是技术人员预先设置多个，再根据 3D 语义地图场景类别等来进行选取，也可以由技术人员预设设置一定的操作教程生成规则，再根据 3D 语义地图场景类别等来操作教程的生成得到。

作为本发明的一个实施例，在进行操作模式确定时，包括：

识别 3D 语义地图对应的场景类别。

基于场景类别及 3D 语义地图中包含的物品，获取对应的操作模式。

在本发明实施例中，考虑到用户在不同实际场景下的不同需求，会同时根据用户 3D 语义地图对应的场景类别和实际包含的物品来判断用户的实际可能需求，例如用户处于厨房之中，且包含大量食物，则说明用户极有可能需要进行烹饪，此时可以生成对应的一些做饭的教程。其中，具体的场景类别识别方法以及物品识别方法此处不予限定，可由技术人员根据实际需求设定。

作为本发明的另一个实施例，在进行操作模式确定时，包括：

识别 3D 语义地图对应的场景类别，并获取用户的用户数据。

基于场景类别、3D 语义地图中包含的物品及用户数据，获取对应的操作模式。

考虑到即使是在相同的场景，具有相同的物品的情况下，不同用户的实际需求还是可能存在一定区别，例如即使都是处于厨房之中且包含大量食物，但每个用户的口味可能不同，其对应的烹饪教程也必然存在较大差别，因此，仅根据用户所处环境的场景类别以及包含的物品情况有时难以实现对用户操作教程需求的准确识别。在本发明实施例中，会同时根据环境的场景类别、实际包含的物品以及用户的一些用户数据来确定出用户的实际操作需求，并筛选出对应的操作教程。其中用户数据的具体内容可以由技术人员或用户自行设定，包括但不限于如用户的个人信息和爱好数据等，或者用户自己预先设置的一些操作需求，例如用户预先设置自己的饮食需求。

S104，监测用户对物品的操作是否满足操作步骤的要求。

在获取到用户所需的操作教程之后，本发明实施例开始对用户的实际操作进行监测识别，并与操作教程之中的操作步骤进行比对，判断用户实际操作是否符合操作教程的要求，从而确定出用户存在的操作缺陷。其中，为了实现对用户操作的监测识别，本发明实施例中，智能眼镜中的摄像头会对用户的行为进行拍摄，并会对得到的图像/视频进行行为分析，即分析用户的行为动作和行为顺序等是否满足操作教程中的要求。

S105，若用户对物品的操作不满足操作步骤的要求，输出操作步骤对应的操作提示。

当用户出现不满足操作教程中操作步骤要求的操作时，说明用户出现了操作错误，因此此时本发明实施例会对用户进行操作提示，即告知用户当前操作存在错误，并会同时告知用户正确的操作步骤，例如，对设备进行操作时，在按下开机键之后，需要使用模式键选取设备模式后，再按开始键进行，而若用户在按下开机键之后，直接就按开始键，此时本发明实施例发现用户操作不满足操作教程的要求，就会提示用户，不应按开始键，而是需要按模式键选取设备模式。其中，操作提示的输出方式包括但不限于如音频/视频/文字等提示，具体可由技术人员自行设定。

作为本发明实施例一中进行提示输出的一种具体实现方式，包括：

识别用户在 3D 语义地图中的注视区域，并基于注视区域进行操作提示的增强现实输出。

本发明实施例中，会利用增强现实（Augmented Reality， AR）格式技术来进行提示信息的输出，以使得用户可更为直观地获知自己的问题以及对应的正确操作。

本发明实施例通过构建用户所处环境的 3D 语义地图，并在识别出用户存在物品注视行为，即确定用户存在操作提示需求时，根据 3D 语义地图实际情况智能识别出用户所需的操作模式（即对物品的操作教程），并根据用户对物品的实际操作情况，来进行监测和提示用户操作，使得用户无需再进行任何手动操作输入，也无需一直主动对着屏幕观看，即可以及时得知自己操作中的问题，实现了对用户操作教程的智能匹配，极大地提升了用户对操作错误的获知效率。

作为本发明实施例一中 3D 语义地图构建的一种具体实现方式，在本发明实施例中，需要采集的环境图像包括环境的彩色图像和深度图像，如图 2 所示，本发明实施例二，包括：

S201，基于彩色图像和深度图像，得到用户的位置信息和姿态信息，以及用户所处环境内物品的位置信息和物品信息。

在本发明实施例中，智能眼镜以 RGB-D 相机作为传感器，获取彩色图像和深度图像，利用视觉 SLAM 算法完成智能眼镜的自主定位和位姿估计及优化（即用户的位置信息和

姿态信息获取），同时进行物品检测获取环境的语义信息，然后采用 RGB-D 分割算法将感知到的物品分割出来，从而构建环境的 3D 语义地图。具体而言，视觉 SLAM 基本框架由视觉里程计、后端优化、回环检测和三维建图四个部分组成。视觉里程计要完成相邻两帧图像间运动的估计，粗略估计出相机当前的位姿。后端优化就是要对视觉里程计的估计进行全局一致的优化，消除噪声干扰，另外利用回环检测的约束优化位姿，使定位和位姿估计更加准确。回环检测是在重新回到原来经过的位置时，消除过程的累计误差。地图的构建基于前三部分估计的运动和位姿，创建环境的三维地图。

视觉 SLAM 算法整体设计，通过传感器输入数据实现对当前位置和姿态的全局一致估计，是对自身的定位和运动的明确。视觉 SLAM 算法由视觉里程计、后端优化和回环检测三个部分协作完成。

视觉里程计通过提取 ORB 特征点，利用 FLANN 算法匹配相邻两帧 ORB 特征，利用匹配结果采用 PnP 算法和 RANSAC 算法结合完成智能眼镜位置和姿态的粗略估计。视觉里程计的任务是估计出智能眼镜在两帧图像之间的位姿变化，估计出智能眼镜一段时间内的姿态和运动轨迹。其由以下进程构成：

1、特征提取

我们采用 ORB 特征提取的方式提取图像特征。ORB 特征由 FAST 关键点和 BRIEF 描述子构成，并人为地赋予了旋转和尺度不变性。

特征提取的过程：

1）粗提取，对于某一点 p，其像素值为 Ip，以 p 为中心半径为 3 的圆上的 16 个点，如果存在 12 个以上的点的像素值与 Ip 相差在阈值内，则认为该 p 是一个候选 FAST 关键点。

2）计算每个子集的信息熵。以信息增益为评价标准，其值最高的像素点设定为决策树的根节点，并继续对其子集进行迭代，直到确定该点的性质，即 FAST 关键点或非 FAST 关键点。于是便生成了 ID3 决策树，利用该树筛选出最优的 FAST 关键点。

3）利用非极大值抑制的思想，在局部范围内，保留得分最高的 FAST 关键点，删掉其他得分较低的 FAST 关键点，遍历一遍，即可完成筛选。

4），赋予特征尺度和旋转不变性。尺度不变性利用金字塔原理实现，将图像降采样处理得到图像金字塔，对其每一层都完成上述四步特征提取，实现 FAST 关键点的尺度不变性。旋转不变性通过灰度质心法实现，计算以关键点为中心图像块 U 的质心，把从中心到质心的向量定义为关键点的方向，实现 FAST 关键点的旋转不变性

描述子：在 BRIEF 基础上做出改进。一是考虑关键点的 31×31 邻域内所有点，

将图像进行高斯平滑滤波后，选取 5×5 邻域内灰度均值代替单个点灰度进行计算，抗噪性强。二是选取 5×5 邻域时采用均值接近 0.5 的不相关贪婪搜索算法，保证了描述子具有代表性和独特性，使其具备区分性。

2、特征匹配

选用快速近似最近邻（FLANN）算法，核心思想是利用导出索引结构将搜索范围定位在特征邻近区域，在邻近区域内完成特征匹配，可以有效地加快匹配速度。利用 BRIEF 描述子由 0 和 1 组成的特点，使用局部敏感哈希作为导出索引结构。将特征按照同样的方式投影到哈希空间，本来相邻的两个特征投影后仍然在哈希空间相邻的几率也会非常高，而之前不相邻的两个特征投影之后在哈希空间相邻的几率会很小，这样即可在哈希空间邻域内进行匹配，有效的缩小了范围。

3、位姿估计算法设计

在提取并匹配两个相邻帧的特征之后，使用匹配关系来估计智能眼镜的运动和姿态。为了定量的估计智能眼镜运动及位姿，首先必须清楚智能眼镜成像与空间点的数学几何关系。智能眼镜成像的过程也被称为观测过程，即三维空间中的点反射或发射光线，经过智能眼镜的光心，投影到智能眼镜的成像平面上的过程。

本发明采用 PnP 算法进行两帧间位姿的粗略估计，再利用 RANSAC 算法进行该位姿的帧间一致性优化，避免误匹配特征对位姿估计造成严重影响。采用 RANSAC 算法对该问题进行优化。

首先提取图像的 ORB 特征点，利用快速最近邻算法进行相邻两帧图像的特征匹配，最后使用 PnP 算法和 RANSAC 算法完成相机运动和姿态估计。

4、后端优化

由于数据噪声、误匹配、计算误差等因素影响，会造成位姿估计的误差，长时间运行时误差逐渐累积，会严重影响系统性能。后端优化就是要对视觉里程计的估计进行全局一致的优化，消除噪声干扰，使位姿估计更加准确。另外，当系统检测到回环后，将信息传递给后端，消除累计误差。

本发明实施例引入关键帧机制，选择有代表性的图像帧进行位姿优化，减少不必要的计算。对于局部位姿优化问题，采用捆集调整算法，优化相机目前处于收敛过程的姿态和特征点。当系统检测出回环时，后端采用位姿图优化方式，获取全局一致的轨迹和位姿。后端优化接收视觉里程计传来的位姿和特征点，采用捆集调整的方式进行优化。工作流程为：检查队列、处理关键帧，地图点剔除、生成和融合，局部捆集调整优化位姿。

满足以下四条原则即将图像设定为关键帧：

（1）由于相机获取数据帧频较高，当前关键帧和上一帧关键帧之间必须经过一定的序列间隔。

（2）后端优化部分不在工作状态下。

（3）当前关键帧与之前选择地所有关键帧相互的共视区域低于一定范围。

（4）当前帧拥有足够的特征点和匹配，保证特征的丰富性。

5、回环检测

回环检测，主要解决智能眼镜位姿估计误差逐渐积累的问题。当相于再次来到曾经到过的地方时，该部分确认并建立当前位姿和历史位姿间的连接关系，传到后端进行优化处理，将系统长时间运行的累积误差消除，得到全局一致的轨迹和位姿。另一方面，回环检测提供了当前数据与所有历史数据的关联，在视觉里程计对特征的跟踪丢失时，可以利用回环检测进行重定位，增强位姿估计的鲁棒性。

回环检测在系统运行时持续检测，通过发生回环的约束来消除智能眼镜位姿估计的累计误差。将相机回到某个曾经来过的位置时的闭环约束关系传到后端进行位姿图优化。回环检测的工作流程包括：步骤 1、回环候选帧检测。步骤 2、与前关键帧建立联系。步骤 3、检测是否发生回环，若未发生，返回值步骤 1，若发生进入步骤 4。步骤 4、位姿图优化。

5.1、磁带模型的建立

把特征看作一个个单词，预先训练包含所有特征类型的字典，对每个图像的特征按照该字典生成一个对应词的集合，也就是词袋。于是，判断图像的相似程度时只要对比它们的词袋即可，大大加快了回环检测的速度。特征聚类使用无监督机器学习 K-means++ 算法，使用 K-d 树的结构提高搜索效率。

字典的训练过程为：

1）在根节点，用上述 K-mean++算法将所有样本分成 k 类，得到第一层。

2）对每一层的每个节点，同样用 K-mean++将属于该节点的样本再聚类为 k 类，得到新的一层。

3）以此类推，直到最后的叶子节点层。该叶子层就是特征对应的单词。

5.2 相似度计算方法

引入 TF-IDF 算法。如果当前帧与之前的某个关键帧的相似度超过当前帧与上一关键帧的相似度的 3 倍，就认为可能发生了回环。但仍需要一个验证步骤，设立回环检测的缓存机制，单次相似度高不足以判断为回环，当连续帧的相似度都很高时，才确认发生回环。确认发生回环后，回环检测部分将此信息发送给后端，后端采用图优化方式对位姿进行优化，消除累计误差，得到全局一致的轨迹和位姿。

视觉 SLAM 算法整体流程的总结：

视觉 SLAM 算法设计由三部分组成：视觉里程计、后端优化、回环检测。视觉里程计部分首先提取传感器输入帧的 ORB 特征，之后使用快速最近邻算法进行相邻两帧图像的特征匹配，采用 PnP 算法和 RANSAC 算法结合完成相机位姿估计。后端部分对视觉里程计传来的位姿和特征点采用捆集调整的方式进行优化。回环检测利用词袋模型和训练好的字典检测相机是否回到之前到过的位置，如果发生回环，将此约束信息发送到后端，后端采用图优化方式对位姿进行优化，消除累计误差，保证了相机轨迹和位姿的全局一致性。

为了实现对物品的物品信息和位置信息获取，本发明实施例中，首先会从彩色图像中进行物品检测，其中检测算法包括但不限于如 Yolv V3 物体检测算法，此处不予限定，并从彩色图像中进行物品分割，详述如下：

在获取彩色图像的同时也获取了相应像素的深度图，也就是三维点云。因此，建立语义地图前需要将检测到的物体分割出来，再利用视觉 SLAM 模块估计并优化出的相机位姿将像素点投影到空间中的位置，从而构建三维的语义地图。我们通过一种改进 GrabCut 算法实现目标 RGB-D 分割，该算法结合 CPF（Constrained Plane Fitting）算法分割的几何平面信息，提升了 GrabCut 的分割效果，实现三维目标的 RGB-D 分割。首先采用 GrabCut 算法分割图像，再利用 CPF 算法分割三维点云，最后将点云分割结果作为滤波器，剔除图像分割结果中不符合目标空间几何关系的像素点，完成目标的 RGB-D 分割。

S202，根据用户的位置信息及姿态信息以及物品的物品信息和位置信息，构建 3D 语义地图。

至此已经获取构建语义地图的全部所需信息，包括优化过的关键帧的智能眼镜位置及姿态（即用户的位置信息和姿态信息）、关键帧中物品的类别（即物品信息）及其位置和检测到的目标物体的三维分割。接下来要做的就是整合获取的信息构建全局一致的三维语义地图，大致过程分为三步：首先进行数据的一致性融合并更新目标物体模型，然后构建环境的三维地图，最后将语义信息融合到三维地图中，得到包含丰富信息的三维语义地图。

语义地图的构建算法研究与设计。首先设计了一种改进 GrabCut 的目标 RGB-D 分割算法，将 GrabCut 分割彩色图与 CPF 分割深度图的信息相结合，将检测到的目标分割出来，完成以对象为单位的语义建图的目标提取工作。并按照物体类别做好标记，再利用目标对象的一致性融合更新物体对象的模型，避免对同一目标的多次建模。进而以彩色八叉树地图结构来构建并存储包含丰富信息的三维语义地图。

1、数据关联与模型更新算法设计

数据关联的作用是得到目标经 RGB-D 分割后的结果时确定该目标是否已经在地图中，

需要添加新的对象还是对现有对象进行维护，避免对同一对象多次建模从而在地图中出现重影。

首先对于每一次检测，基于分割后点云的各质心的欧氏距离选择一组候选对象的边界标志。然后对地图现有边界标志和当前目标的三维点进行最近邻搜索，并计算相关点对的欧氏距离。两个三维点的欧氏距离也就是两点的 2-范数。

如果目标中有一半以上的三维点都与地图现有的目标距离小于一定阈值时，认为该目标与地图现有目标是同一个，将当前目标信息与地图的现有目标相关联，以共同维护目标模型。另外，当所处环境较复杂时，三维点的最近邻搜索通过寻求 k-d 树结构来加速。为了保证数据关联能获取到最新的信息，只要检测到物体就利用上文设计改进的 RGB-D 目标分割算法将其分割出来。这样地图中的每个对象都保留三项信息：通过数据关联得到的目标模型、观测到该目标的关键帧位姿和目标检测模块给出的各类别概率。地图中目标物体的概率可以根据目标检测模块给出的概率值更新，设当前共检测 C 类物体，Sc 表示该目标的各类别概率组成的向量，n 为检测到此目标的关键帧个数，目标检测概率更新计算：

$$Sc = \sum_{i=1}^{n} Si \quad (1)$$

则地图中该目标的所属类别为 max(Sc)，其置信度为 p= max(Sc)ln，为语义地图中目标类别及概率的标注提供信息。

2、语义地图的构建与存储形式

本发明采用灵活、占内存量小又支持实时更新的地图形式：八叉树地图

整个空间作为根节点，按照空间坐标系分为八个子节点，再继续将每个子节点分为八个子节点，一直分到所需的分辨率，即叶子节点。八叉树地图与点云图的体素模型不同点在于，当某方块中所有点都被占据或不占据时，就没有必要展开这个节点，相对来说占用内存空间非常小。而且搜索叶子节点的速度非常快，d 层的八叉树时间复杂度为 O(d)。

另外，八叉树地图能够支持设置每个节点的颜色，也就是彩色八叉树地图形式，同时支持随时更新和修正信息，非常适用于构建三维语义地图。因此，本文选择彩色八叉树地图作为三维语义地图的构建与存储形式。

构建三维语义地图，首先建立环境的三维地图并持续更新，再实时地将语义信息融合到三维地图中，即可以构建包含丰富信息的三维的语义地图。在相机运动过程中不断地获取信息并处理，进而不断地更新语义地图。

三维建图的工作是根据视觉 SLAM 算法中对关键帧估计并优化后的位置和姿态，将 RGB-D 相机捕捉到的深度信息映射到三维空间中，建立环境的三维地图。由于本发明采用

的 RGB-D 相机可以获取视野中每个像素的深度，可以直接使用深度图进行稠密建图，根据优化的相机位姿，将深度图映射为点云后进行拼接，得到三维地图。

语义地图就是包含语义信息的三维地图，也就是在建立八叉树地图时标注出环境的语义信息。在建立三维八叉树地图时，同时将语义信息融入八叉树结构中，就得到包含丰富信息的三维语义地图。

本发明语义建图系统以 RGB-D 相机作为视觉传感器，捕捉彩色信息和深度信息。利用视觉 SLAM 算法完成 AR 智能眼镜的自主定位和位姿估计及优化，得到全局一致的轨迹和位姿。同时利用卷积神经网络 YOLOv3 模型进行目标检测，检测关键帧中出现的物体类别、概率及其位置，获取环境的语义信息。然后采用 RGB-D 分割算法将感知到的物体分割出来，选择八叉树地图形式构建环境的三维地图。最后，将语义信息融入该八叉树地图，完成环境的三维语义地图的构建。

本发明实施例设计了以 RGB-D 相机为传感器的视觉 SLAM 算法。本发明选择 ORB 特征作为算法的基础，不仅在视觉里程计的位姿估计中提供空间几何关系，还作为回环检测时图像相似度判断的标准，一定程度上实现了系统的统一性。本发明实施例设计的位姿估计算法中，采用 PnP 算法替代 ICP 算法，PnP 算法利用前一帧优化好的相机坐标和当前帧像素坐标计算位姿，避免相机测量误差的干扰。后端优化部分采用针对性的算法处理，对视觉里程计传来的位姿和地图点的优化采用局部捆集调整算法，对回环检测传来的闭环约束优化时采用图优化的方式优化位姿。

设计了语义建图系统的整体结构。本文设计了一种改进 GrabCut 的目标 RGB-D 分割算法，结合利用深度点云信息的 CPF 分割结果修正仅利用彩色图像信息的 GrabCut 分割，实现二者性能的互补。将目标检测的对象分割出来，并按照物体类别做好标记，综合视觉 SLAM 算法获取的相机位姿和目标的检测和分割结果，以彩色八叉树地图结构构建并存储环境的三维语义地图。在实验室环境下运行系统，系统在自身定位、姿态估计、语义感知的同时，构建出可读且准确的三维语义地图，验证了本发明语义建图方案的可行性。

作为本发明实施例一中对用户进行注视行为识别的一种具体实现方式，如图 3A 所示，本发明实施例三，包括：

S301，获取用户的眼部图像，对眼部图像进行瞳孔定位，并基于得到的瞳孔位置信息确定用户在 3D 语义地图中的注视区域。

本发明实施例中，会基于用户的瞳孔和普尔钦斑来进行用户视线的追踪，并确定出用户注视的区域，因此首先需要确定出眼睛中瞳孔的位置。其中，具体的瞳孔识别算法可由技术人员根据需求自行设定，包括但不限于如使用神经网络模型来进行瞳孔图像的样本数

据训练，并识别眼部图像中的瞳孔，或者参考本发明实施例四来进行识别。由于基于瞳孔和普尔钦斑来进行用户视线的追踪的技术已较为成熟，因此此处不予赘述。

S302，识别注视区域内包含的物品，并统计注视区域内包含的物品在注视区域内连续存在时长。

S303，若连续存在时长中存在大于预设时长阈值的时长，判定用户注视 3D 语义地图中包含的物品。

在确定出用户的注视区域之后，还需要进一步地确定出注视区域中包含的物品，以及每个物品在注视区域存在的连续时长（即用户对物品的连续注视时长），若该时长较长，则说明用户存在注视某一物品的行为。其中，预设时长阈值的具体值大小，可由技术人员自行设定。

作为本发明的一个实施例，考虑到用户头部会运动，因此在进行上述视线追踪时，视线会发生变化，为了实现更为准确的视线追踪，本发明实施例中包括：

头部的自然运动包括两项基本运动:垂直方向的俯仰运动和水平方向的左右运动。一般的基于二次多项式的映射模型是用户在保持头部静止不动的情况下通过多点校准得到，在头部位置发生变化时，该模型得到的估计注视点误差大大增加。本文在此基础上提出一种基于多项式映射的头动解决算法，该算法需要利用头部运动估计设备实时得到头部运动的信息，在头动的情况下首先利用校准时得到的多项式映射模型进行注视点位置初步估计，将该估计点坐标结合初始的眼睛位置建立三维视线方向，再利用头部运动信息对该视线方向进行旋转和平移补偿，将当前补偿后的视线方向与屏幕的交点作为最终的注视点估计坐标。另外，考虑到其实头部运动的影响是源于头部运动使得眼睛的位置发生了变化，所以我们只要知道的眼睛位置的变化情况就可以对视线进行补偿。

根据以上几点说明建立头部自然运动下的视线估计模型，如图 3B 所示，图中表示一个视线估计原理的侧平面图，O1 表示初始的眼睛位置，O2 表示运动后眼睛的位置。在初始位置当注视屏幕上点 S1 时，对应得到瞳孔角膜矢量 pccr1，头部向右绕 Y 轴旋转 α 角度后，眼睛移到位置 O2，假设 pccr2 等于 pccr1，即眼图特征不发生任何变化，则对应的视线将向右发生偏转，与屏幕的交点我们假设为 S2 ，这是此时的注视点位置，但由于瞳孔角膜矢量未发生变化，用 pccr2 进行视线估计时误差较大。对此我们将该点作为视线估计的初步估计点，将该点和眼睛的初始位置建立视线方向 g1，再利用头部运动的角度信息对 g1 进行修正，将修正后的视线方向与屏幕的交点即为当前的注视点估计点。

作为本发明实施例三中对瞳孔进行定位的一种实现方式，如图 4A 所示，本发明实施例四，包括：

S401，将眼部图像划分为N×M个区域图像，并对所有区域图像灰度二值化处理，得到对应的N×M个眼部灰度值，其中，N和M均为正整数。

如图4B所示，本发明实施例中的眼部特征详述如下，图中基本的矩形块大小一致，图中ABCD是最原始的矩形特征，E是由3个基本矩形组成，F由9个矩形组成，G是一个矩形，H与I由4 个基本矩形构成，J由12个矩形构成，K和L由4个矩形构成，各矩形特征的计算都是图中的黑色部分像素和减去白色部分像素和，这里的特征G是一个单一的矩形特征，所以只计算矩形内的像素和。眼部特征是基于眼睛结构而设计的，由于眼角与周围的亮度有所变化，眼角相对于周围的像素较暗，特征F可以很好的表现这个特点，眼球的中心点基本上是呈现黑色，所以简单的矩形G就表示了眼球的这一特性，眼睛的水平方向会发生两次较明显的像素突变，从巩膜到虹膜，再由虹膜到巩膜，特征H就可以反应出这种变化，同理眼睛的垂直方向上也有类似的灰度变化特点，I特征就是基于这一特点而产生的，它们与C和E合起来能够增强对眼睛水平和垂直灰度变化特点的描述，J表示了眼角与眼球间部分的灰度变化情况，K、L表现了眼角的边缘信息。增加新特征后使得用于分类的眼睛特征数量减少，使眼睛检测变得更加容易。

本发明实施例的瞳孔定位采用了帧独立的设计思想。仔细观察人眼截图发现，人眼图像信号的很多部分会随着入射光线、不同人、眼睛的镜面反射而改变，而且，随着人脸转角的不同，出现在截图中的相对位置也不同，直接进行分类训练效果不尽如人意。经过多次对比，我们发现瞳孔区域是人眼截图中比较稳定的图像信息，在睁眼时候这部分的特征明显。因此我们先定位瞳孔进而简化问题的复杂性。

本发明实施例中，N和M均为正整数，其具体值大小由技术人员自行设定，以M=N=10为例进行说明，此时会将眼部图像划分为10×10个区域图像，并会计算每个区域图像的灰度值，得到10×10个眼部灰度值。

S402，获取用户的皮肤图像，并计算皮肤图像灰度二值化处理后的平均灰度值。

为了实现肤色灰度与最低灰度之间的权衡，本发明实施例还会采集用户的皮肤图像，并计算其对应的平均灰度值。其中皮肤图像优选地，可以为眼部周围的皮肤图像。

S403，按照差值绝对值从小到大的顺序，从对应的眼部灰度值与平均灰度值的差值绝对值小于预设灰度阈值的区域图像中进行图像筛选，直至筛选得到的区域图像包含的像素点数处于预设数量阈值范围内，得到瞳孔对应的区域图像，以确定瞳孔在眼部图像中的位置。

在得到每个区域图像分别对应的眼部灰度值以及计算得到的皮肤的平均灰度值之后，本发明实施例会分别计算每个区域图像的眼部灰度值，与皮肤的平均灰度值的差值绝对值，

并会按照从小到大的顺序依次筛选出符合预设灰度阈值要求的区域图像，并在每次筛选出一个区域图像之后，统计一次所有筛选出的区域图像包含的像素点数，直至筛选出的区域图像包含的像素点数处于预设数量范围内位置，以保证对瞳孔识别的准确可靠。其中，预设灰度阈值和预设数量阈值的具体值大小，可有技术人员自行设定。

作为本发明实施例一中对 3D 语义地图进行场景识别的一种具体实现方法，本发明实施例包括：

为了实现对每张图片进行单独分类，本文使用了 Places205-AlexNet 网络模型，该模型在各种基准数据集上的性能表现均超过其他方法。Places205-AlexNet 网络模型遵循与 AlexNet 相同的网络架构但其在场景分类任务上进行了针对性的训练。该训练数据集包含了 205 个语义类的约 250 万张图片，且每一类至少有 5000 张样本图片。这些图片来源于各种网络资源，如谷歌图片、必应（Bing）、Flickr 等，并将图片进行分类标注。训练数据集的样本数量和种类确保了由此产生的分类器可很好地泛化，且在其从未训练过的环境中应用时不必进行二次训练或微调。这确保了本文的语义建图系统是可移植的、能被在多种环境下作业的用户使用。

Places205-AlexNet 网络模型的输入是被调整为 224×224×3 像素的 RGB 图片，而与它们的初始尺寸无关。Places205-AlexNet 卷积神经网络模型共有 8 层网络，其中包括前面 5 个卷积层和后面 3 个全连接层。给定当前场景的图片 It，该网络的输出层 soft-max 输出在 205 个已知场景类上的离散概率分$p(o_t|I_t)$。本文的分类器使用 Places205-AlexNet 网络模型的 fc7 层的输出作为一个特征向量，fc7 层是网络中最后一个通用的（即与类无关的）全连接层。由于 Places205-AlexNet 网络模型设计的目的是为了对 Places205 场景数据集中 205 个场景类别进行识别，因此最后的 fc8 层和 prob 层具有 205 个输出神经元节点。

假设给定当前图片 It，该网络的输出层 prob 输出在 205 个已知类 Ci 上的离散概率分布$p(o_t|I_t)$。令$\Theta=(o_1,o_2,\ldots,o_N)$表示已知场景类标签的组合向量。然后定义相应的组合似然为：

$$L(Q_t|\Theta_t)=(p(o_0|I_t),p(o_1|I_t),\ldots,p(o_{N-1}|I_t)p(o_N|I_t)) \quad (2)$$

其中表示 t 时刻的图像 It 属于场景类 Ci 的概率，且 Ci 之间相互独立。由于相机获取的相邻两张图片在时间上是连续的，故可以应用递归的贝叶斯滤波技术。本文将机器人的场景分类问题描述为一个概率估计问题并估计在所有可能的场景标签 Ci 上的离散概率分布其中所有从过去到现在的观测图片是已知的。假设满足一阶马尔科夫性质，则可得到如下

的贝叶斯滤波公式：

$$p(\delta_t|I_t) = L(I_t|\delta_t)\, p(\delta_{t-1}|I_{t-1}) \quad (3)$$

在将场景分类问题视为一个贝叶斯估计问题后，可以很自然地整合其他信息资源。例如，由于室内服务机器人工作在室内，因此 205 类场景中的室外场景基本上不可能被观测到。

作为本发明实施例五，考虑到实际情况中，满足场景类型、物品情况以及用户数据要求的操作教程有时可能有多个，此时，为了确保最终确定的操作教程是用户实际所需，如图 5 所示，本发明实施例包括：

S501，获取多种操作模式。

S502，获取用户对的模式选取指令，并基于模式选取指令，从多种操作模式筛选出一种操作模式。

当有多种满足要求的操作教程时，本发明实施例可由用户自行选取其所需的一个进行后续操作。其中，模式选取指令的输入方式，包括但不限于如用户通过语音/眼动/头部姿态变换等方式来进行输入，例如在智能眼镜中显示出多种操作教程的相关信息，用户通过语音选取、眼动操作（如眨眼、注视等）或者头部姿态变化（如摇头切换操作教程，点头确认，以实现模式选取指令输入），具体需由技术人员根据实际需求设定可进行指令交互的方式。

作为本发明实施例中用户与智能眼镜的一直眼动信息交互实现方式，本发明实施例中用户可以进行多种眼动操作，以与智能眼镜进行信息交互操作，包括：

1）判断是否是阅读行为：

用来标注阅读行为（已确认注视目标，不需要更多的识别时间）的主要注视点，其时间长度预设范围内，预设范围优选为 600-1100 毫秒。

2）判断用户回视行为：

借助扫视数据可以定义回视，即注视点坐标落在前 5 个注视点为圆心，1 个注视区域为半径构成的空间中，但该注视点的前 1 个注视点不算。

3）判断是否是改变全局关注点的行为：

整体关注点改变是在上一个主要注视点（持续时间预设时长以上，优选为 1100 毫秒以上）后扫视长度超过 3 个注视区域及以上。

4）判断局部关注点改变的行为：

局部关注点改变是在上一个主要注视点（持续时间预设时长以上，优选为 1100 毫秒以

上）后多个扫视长度不超过 3 个注视区域，但其位置改变总距离超过了 3 个注视区域的行为。

5）对搜索行为的判断：

搜索行为：当出现整体关注点改变或局部关注点改变行为时，可以认为用户正在进行搜索。整体关注点改变是典型的搜索行为，而局部关注点改变则是用户认为大致区域已经找到，正在继续寻找特定目标。

持久搜索：在上一次主要注视点（持续时间预设时长以上，优选为 1100 毫秒以上）后 10 个注视区域及以上的扫视连续出现时的行为，可以认为用户在思考、走神或休息。

本发明实施例中，用户可以通过上述方法来对智能眼镜进行眼动控制，以实现本发明其他实施例中所需的各类人机交互操作。

作为本发明实施例一中进行用户操作监测的一种具体实现方法，如图 6 所示，本发明实施例六，包括：

S601，识别用户对物品操作时的行为，并判断行为是否满足操作步骤的要求。

S602，获取操作模式涉及的物品对应的预设属性阈值，并识别用户对物品进行操作过程中，物品的物品属性数据是否满足预设属性阈值。

S603，若行为不满足操作步骤的要求，和/或物品属性数据不满足预设属性阈值，判定用户对物品的操作不满足操作步骤的要求。

考虑到即使用户操作行为上看似无误，但其中具体每一步操作是否完全准确有时也是难以识别的，例如用户的确是先开机-再选模式-再开始，但在模式选取时选错了模式，此时仍会导致整个操作出现问题，因此，为了及时确定出用户的操作问题，本发明实施例对用户进行操作监测时，不仅仅只是监测用户的操作行为，同时还会对用户操作物品的属性据进行监测，以防止此类错误发生了却又无法识别的情况出现。

具体而言，本发明实施例中会对操作教程内涉及的每个物品均设置对应的属性阈值，例如某个做饭教程中，设定烤箱的烤制时间为 1 小时温度为 120 度，并对用户操作过程中物品的属性数据进行识别，判断其操作导致的物品属性数据变化是否处于阈值要求范围，如上述的，判断用户设定的烤箱的烤制时间和温度是否为 1 小时和 120 度（可根据智能眼镜拍摄到的用户设置烤箱时的图像来进行识别得到）。

当用户行为出现错误（如按错键），或者用户操作导致的物品的属性数据不满足要求（如烤箱温度设置错误）时都说明用户的操作是错误的，因此本发明实施例都会判定用户对物品的操作不满足操作教程中草在步骤的要求。

对应于上文实施例的方法，图 7 示出了本发明实施例提供的操作提示装置的结构框图，

为了便于说明，仅示出了与本发明实施例相关的部分。图 7 示例的操作提示装置可以是前述实施例一提供的操作提示方法的执行主体。

参照图 7，该操作提示装置包括：

地图构建模块 71，用于获取用户所处环境的图像，并基于所述图像构建用户所处环境的 3D 语义地图。

注视识别模块 72，用于对用户进行眼动识别，判断所述用户是否注视所述 3D 语义地图中包含的物品。

模式获取模块 73，用于若所述用户注视所述 3D 语义地图中包含的物品，获取所述 3D 语义地图对应的操作模式，所述操作模式中包含对一个或多个所述物品的操作步骤。

操作监测模块 74，用于监测所述用户对所述物品的操作是否满足所述操作步骤的要求。

操作提示模块 75，用于若所述用户对所述物品的操作不满足所述操作步骤的要求，输出所述操作步骤对应的操作提示。

进一步地，地图构建模块 71，包括：

基于所述彩色图像和所述深度图像，得到所述用户的位置信息和姿态信息，以及所述用户所处环境内所述物品的位置信息和物品信息。

根据所述用户的位置信息及姿态信息以及所述物品的物品信息和位置信息，构建所述 3D 语义地图。

进一步地，注视识别模块 72，包括：

瞳孔定位模块，用于获取所述用户的眼部图像，对所述眼部图像进行瞳孔定位，并基于得到的瞳孔位置信息确定所述用户在所述 3D 语义地图中的注视区域。

时长统计模块，用于识别所述注视区域内包含的物品，并统计所述注视区域内包含的物品在所述注视区域内连续存在时长。

注视确认模块，用于若所述连续存在时长中存在大于预设时长阈值的时长，判定所述用户注视所述 3D 语义地图中包含的物品。

进一步地，瞳孔定位模块，包括：

将所述眼部图像划分为 N×M 个区域图像，并对所有所述区域图像灰度二值化处理，得到对应的 N×M 个眼部灰度值，其中，N 和 M 均为正整数。

获取所述用户的皮肤图像，并计算所述皮肤图像灰度二值化处理后的平均灰度值。

按照差值绝对值从小到大的顺序，从对应的所述眼部灰度值与所述平均灰度值的差值绝对值小于预设灰度阈值的所述区域图像中进行图像筛选，直至筛选得到的所述区域图像

包含的像素点数处于预设数量阈值范围内，得到瞳孔对应的所述区域图像，以确定所述瞳孔在所述眼部图像中的位置。

进一步地，模式获取模块 73，包括：

识别所述 3D 语义地图对应的场景类别。

基于所述场景类别及所述 3D 语义地图中包含的物品，获取对应的所述操作模式。

进一步地，模式获取模块 73，还包括：

场景识别模块，用于识别所述 3D 语义地图对应的场景类别，并获取所述用户的用户数据。

模式确定模块，用于基于所述场景类别、所述 3D 语义地图中包含的物品及所述用户数据，获取对应的所述操作模式。

进一步地，模式确定模块，包括：

获取多种所述操作模式。

获取所述用户对的模式选取指令，并基于所述模式选取指令，从所述多种所述操作模式筛选出一种所述操作模式。

进一步地，操作监测模块 74，包括：

识别所述用户对所述物品操作时的行为，并判断所述行为是否满足所述操作步骤的要求。

获取所述操作模式涉及的所述物品对应的预设属性阈值，并识别所述用户对所述物品进行操作过程中，所述物品的物品属性数据是否满足所述预设属性阈值。

若所述行为不满足所述操作步骤的要求，和/或所述物品属性数据不满足所述预设属性阈值，判定所述用户对所述物品的操作不满足所述操作步骤的要求。

进一步地，操作提示模块 75，包括：

识别所述用户在所述 3D 语义地图中的注视区域，并基于所述注视区域进行所述操作提示的增强现实输出。

本发明实施例提供的操作提示装置中各模块实现各自功能的过程，具体可参考前述图 1 所示实施例一的描述，此处不再赘述。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本发明实施例的实施过程构成任何限定。

还应理解的是，虽然术语“第一”、“第二”等在文本中在一些本发明实施例中用来描述各种元素，但是这些元素不应该受到这些术语的限制。这些术语只是用来将一个元素与另一元素区分开。例如，第一表格可以被命名为第二表格，并且类似地，第二表格可以

被命名为第一表格，而不背离各种所描述的实施例的范围。第一表格和第二表格都是表格，但是它们不是同一表格。

图 8 是本发明一实施例提供的眼镜的示意图。如图 8 所示，该实施例的眼镜 8 包括：处理器 80、存储器 81，所述存储器 81 中存储有可在所述处理器 80 上运行的计算机程序 82。所述处理器 80 执行所述计算机程序 82 时实现上述各个操作提示方法实施例中的步骤，例如图 1 所示的步骤 101 至 105。或者，所述处理器 80 执行所述计算机程序 82 时实现上述各装置实施例中各模块/单元的功能，例如图 7 所示模块 71 至 75 的功能。

所称处理器 70 可以是中央处理单元（Central Processing Unit，CPU），还可以是其他通用处理器、数字信号处理器（Digital Signal Processor，DSP）、专用集成电路（Application Specific Integrated Circuit，ASIC）、现成可编程门阵列（Field-Programmable Gate Array，FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

所述存储器 71 可以是所述眼镜 7 的内部存储单元，例如眼镜 7 的硬盘或内存。所述存储器 71 也可以是所述眼镜 7 的外部存储设备，例如所述眼镜 7 上配备的插接式硬盘，智能存储卡（Smart Media Card，SMC），安全数字（Secure Digital，SD）卡，闪存卡（Flash Card）等。进一步地，所述存储器 71 还可以既包括所述眼镜 7 的内部存储单元也包括外部存储设备。所述存储器 71 用于存储所述计算机程序以及所述眼镜所需的其他程序和数据。所述存储器 71 还可以用于暂时地存储已经发送或者将要发送的数据。

另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

所述集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本发明实现上述实施例方法中的全部或部分流程，也可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一计算机可读存储介质中，该计算机程序在被处理器执行时，可实现上述各个方法实施例的步骤。其中，所述计算机程序包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括：能够携带所述计算机程序代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器（Read-Only Memory，ROM）、随机存取存储器（Random Access Memory，RAM）、电载波信号、电信信号以及软件分发介质等。

以上所述实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述实施例对本

发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使对应技术方案的本质脱离本发明各实施例技术方案的精神和范围，均应包含在本发明的保护范围之内。

权利要求书

1. 一种操作提示方法，其特征在于，包括：

获取用户所处环境的图像，并基于所述图像构建用户所处环境的 3D 语义地图；

对用户进行眼动识别，判断所述用户是否注视所述 3D 语义地图中包含的物品；

若所述用户注视所述 3D 语义地图中包含的物品，获取所述 3D 语义地图对应的操作模式，所述操作模式中包含对一个或多个所述物品的操作步骤；

监测所述用户对所述物品的操作是否满足所述操作步骤的要求；

若所述用户对所述物品的操作不满足所述操作步骤的要求，输出所述操作步骤对应的操作提示。

2. 如权利要求 1 所述的操作提示方法，其特征在于，所述图像包括彩色图像和深度图像，所述获取用户所处环境的图像，并基于所述图像构建构建用户所处环境的 3D 语义地图，包括：

基于所述彩色图像和所述深度图像，得到所述用户的位置信息和姿态信息，以及所述用户所处环境内所述物品的位置信息和物品信息；

根据所述用户的位置信息及姿态信息以及所述物品的物品信息和位置信息，构建所述 3D 语义地图。

3. 如权利要求 1 所述的操作提示方法，其特征在于，所述对用户进行眼动识别，判断所述用户是否注视所述 3D 语义地图中包含的物品，包括：

获取所述用户的眼部图像，对所述眼部图像进行瞳孔定位，并基于得到的瞳孔位置信息确定所述用户在所述 3D 语义地图中的注视区域；

识别所述注视区域内包含的物品，并统计所述注视区域内包含的物品在所述注视区域内连续存在时长；

若所述连续存在时长中存在大于预设时长阈值的时长，判定所述用户注视所述 3D 语义地图中包含的物品。

4. 如权利要求 3 所述的操作提示方法，其特征在于，所述对所述眼部图像进行瞳孔定位，包括：

将所述眼部图像划分为 N×M 个区域图像，并对所有所述区域图像灰度二值化处理，得到对应的 N×M 个眼部灰度值，其中，N 和 M 均为正整数；

获取所述用户的皮肤图像，并计算所述皮肤图像灰度二值化处理后的平均灰度值；

按照差值绝对值从小到大的顺序，从对应的所述眼部灰度值与所述平均灰度值的差值绝对值小于预设灰度阈值的所述区域图像中进行图像筛选，直至筛选得到的所述区域图像

包含的像素点数处于预设数量阈值范围内，得到瞳孔对应的所述区域图像，以确定所述瞳孔在所述眼部图像中的位置。

5. 如权利要求 1 所述的操作提示方法，其特征在于，所述获取所述 3D 语义地图对应的操作模式，包括：

识别所述 3D 语义地图对应的场景类别；

基于所述场景类别及所述 3D 语义地图中包含的物品，获取对应的所述操作模式。

6. 如权利要求 1 所述的操作提示方法，其特征在于，所述获取所述 3D 语义地图对应的操作模式，包括：

识别所述 3D 语义地图对应的场景类别，并获取所述用户的用户数据；

基于所述场景类别、所述 3D 语义地图中包含的物品及所述用户数据，获取对应的所述操作模式。

7. 如权利要求 6 所述的操作提示方法，其特征在于，获取所述操作模式的过程包括，包括：

获取多种所述操作模式；

获取所述用户对的模式选取指令，并基于所述模式选取指令，从所述多种所述操作模式筛选出一种所述操作模式。

8. 如权利要求 1 所述的操作提示方法，其特征在于，所述监测所述用户对所述物品的操作是否满足所述操作步骤的要求，包括：

识别所述用户对所述物品操作时的行为，并判断所述行为是否满足所述操作步骤的要求；

获取所述操作模式涉及的所述物品对应的预设属性阈值，并识别所述用户对所述物品进行操作过程中，所述物品的物品属性数据是否满足所述预设属性阈值；

若所述行为不满足所述操作步骤的要求，和/或所述物品属性数据不满足所述预设属性阈值，判定所述用户对所述物品的操作不满足所述操作步骤的要求。

9. 如权利要求 1 所述的操作提示方法，其特征在于，所述输出所述操作步骤对应的操作提示，包括：

识别所述用户在所述 3D 语义地图中的注视区域，并基于所述注视区域进行所述操作提示的增强现实输出。

10. 一种眼镜，其特征在于，所述眼镜包括存储器、处理器，所述存储器上存储有可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现如权利要求 1

至 9 任意一项所述方法的步骤。

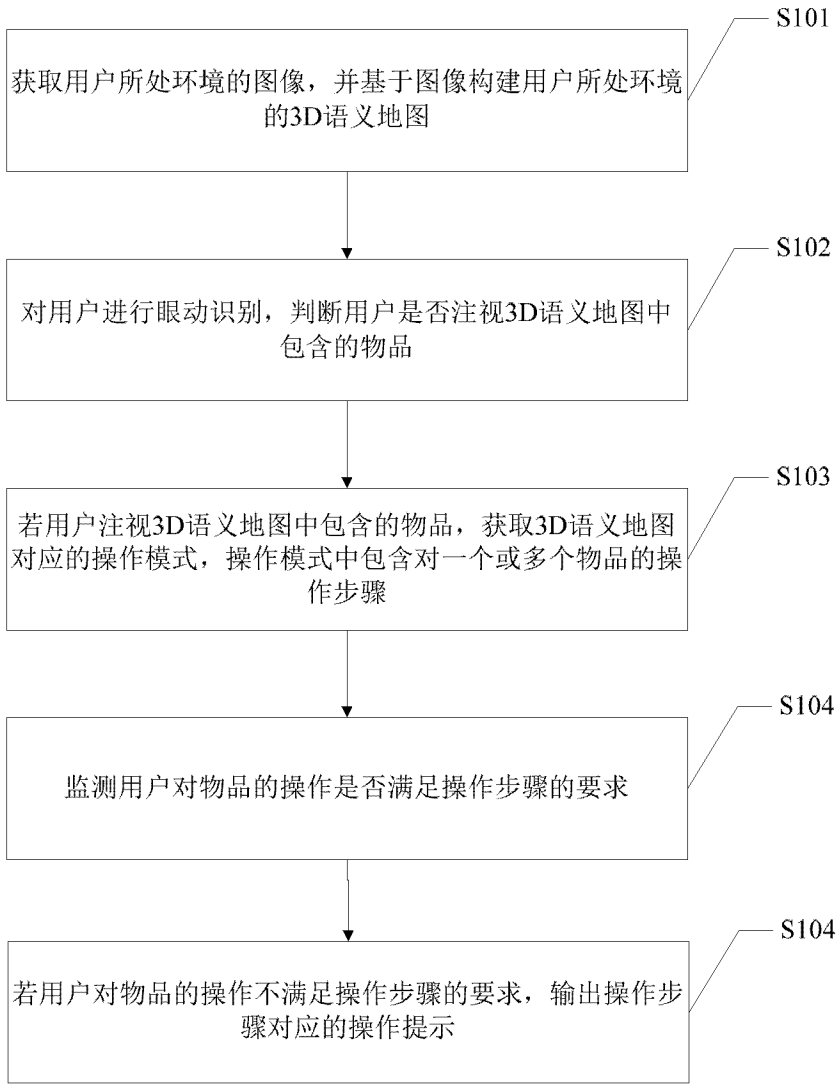

S101
获取用户所处环境的图像，并基于图像构建用户所处环境的3D语义地图
S102
对用户进行眼动识别，判断用户是否注视3D语义地图中包含的物品
S103
若用户注视3D语义地图中包含的物品，获取3D语义地图对应的操作模式，操作模式中包含对一个或多个物品的操作步骤
S104
监测用户对物品的操作是否满足操作步骤的要求
S104
若用户对物品的操作不满足操作步骤的要求，输出操作步骤对应的操作提示

图 1

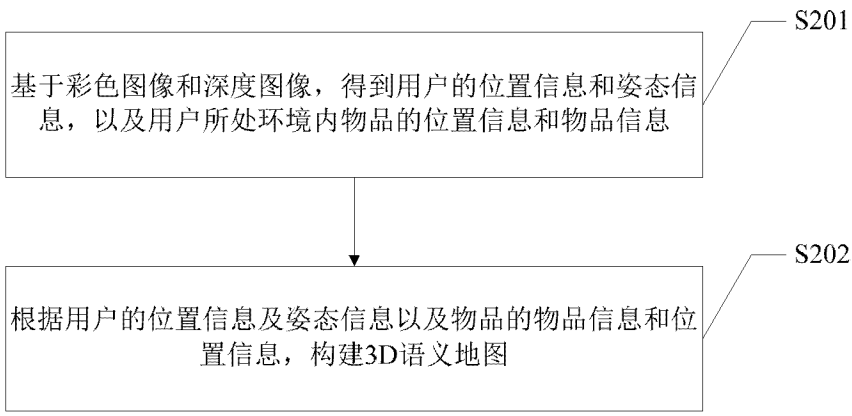

S201
基于彩色图像和深度图像，得到用户的位置信息和姿态信息，以及用户所处环境内物品的位置信息和物品信息
S202
根据用户的位置信息及姿态信息以及物品的物品信息和位置信息，构建3D语义地图

图 2

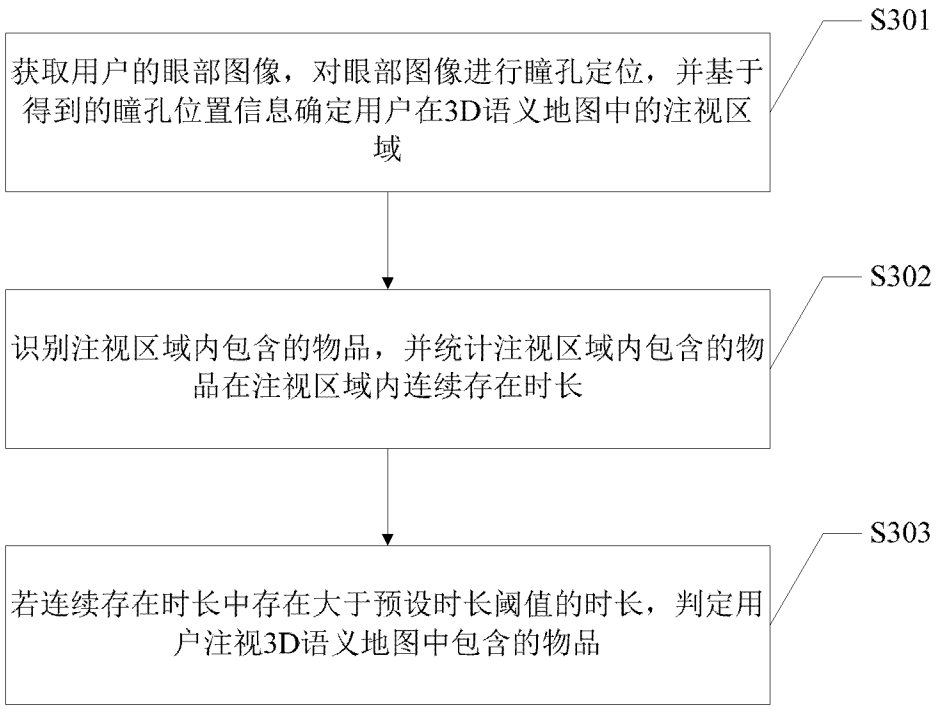

S301
获取用户的眼部图像，对眼部图像进行瞳孔定位，并基于得到的瞳孔位置信息确定用户在3D语义地图中的注视区域
S302
识别注视区域内包含的物品，并统计注视区域内包含的物品在注视区域内连续存在时长
S303
若连续存在时长中存在大于预设时长阈值的时长，判定用户注视3D语义地图中包含的物品

图 3A

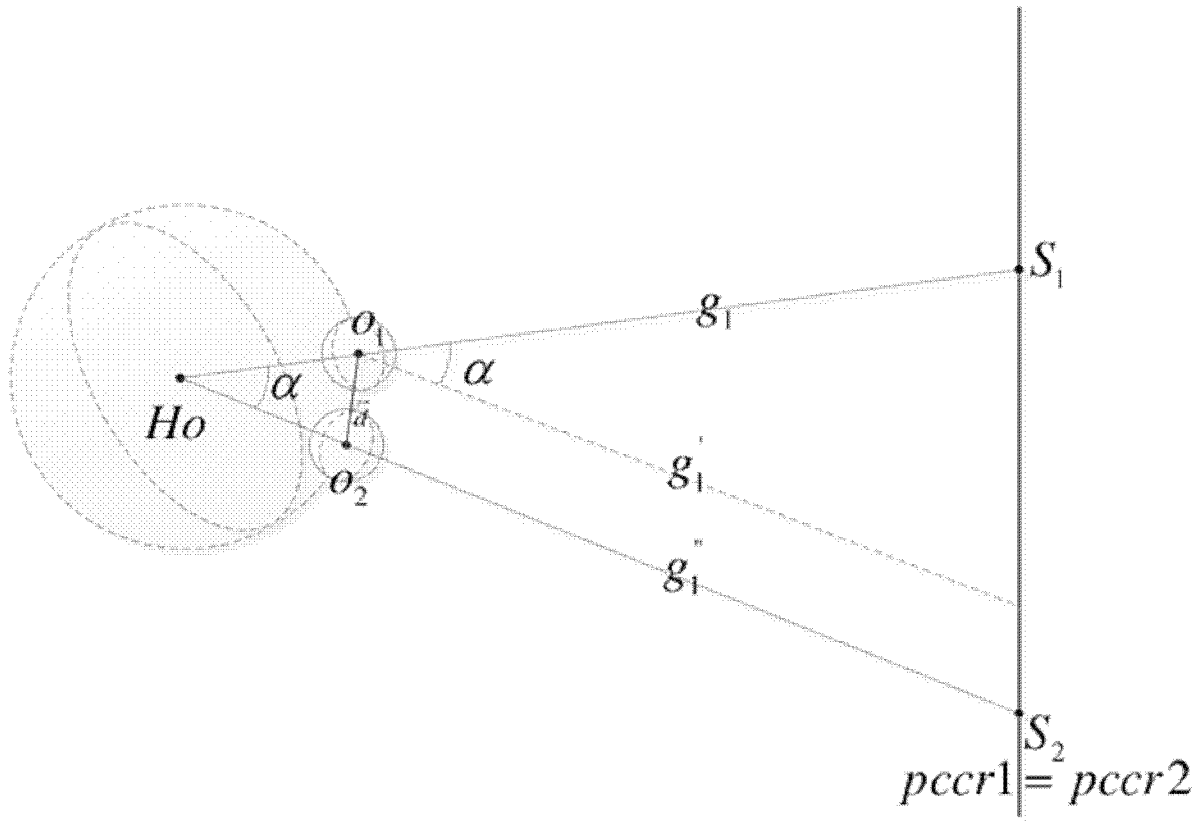

S_1
g_1
o_1
α
α
Ho
d
o_2
g_1'
g_1''
S_2
$pccr1 = pccr2$

图 3B

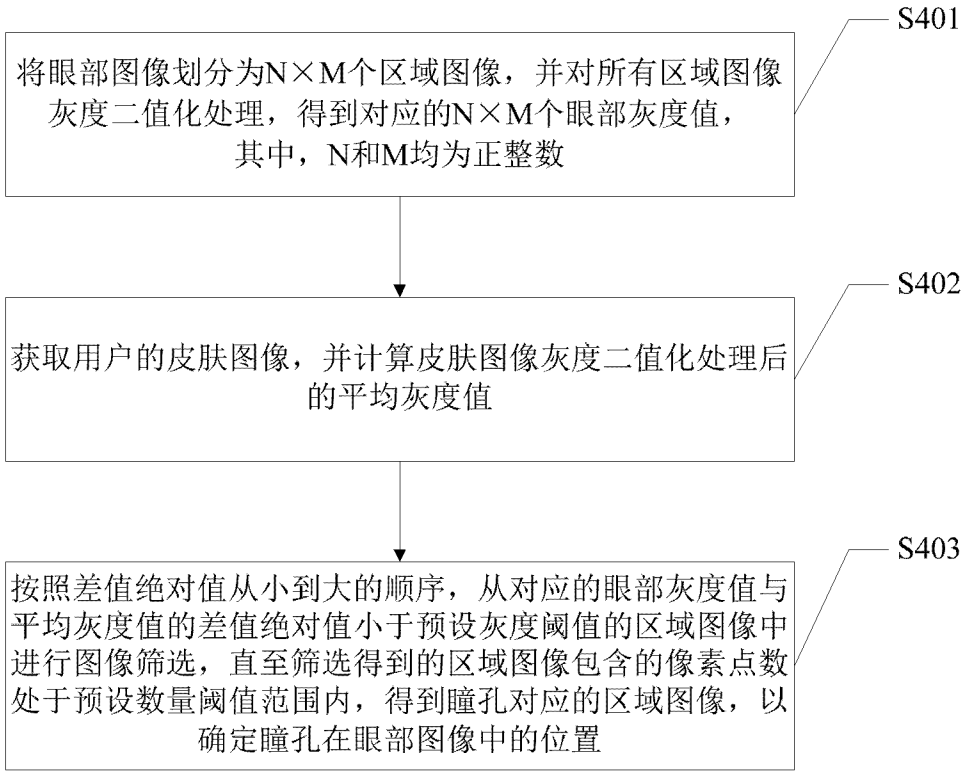

S401
将眼部图像划分为N×M个区域图像，并对所有区域图像灰度二值化处理，得到对应的N×M个眼部灰度值，其中，N和M均为正整数
S402
获取用户的皮肤图像，并计算皮肤图像灰度二值化处理后的平均灰度值
S403
按照差值绝对值从小到大的顺序，从对应的眼部灰度值与平均灰度值的差值绝对值小于预设灰度阈值的区域图像中进行图像筛选，直至筛选得到的区域图像包含的像素点数处于预设数量阈值范围内，得到瞳孔对应的区域图像，以确定瞳孔在眼部图像中的位置

图 4A

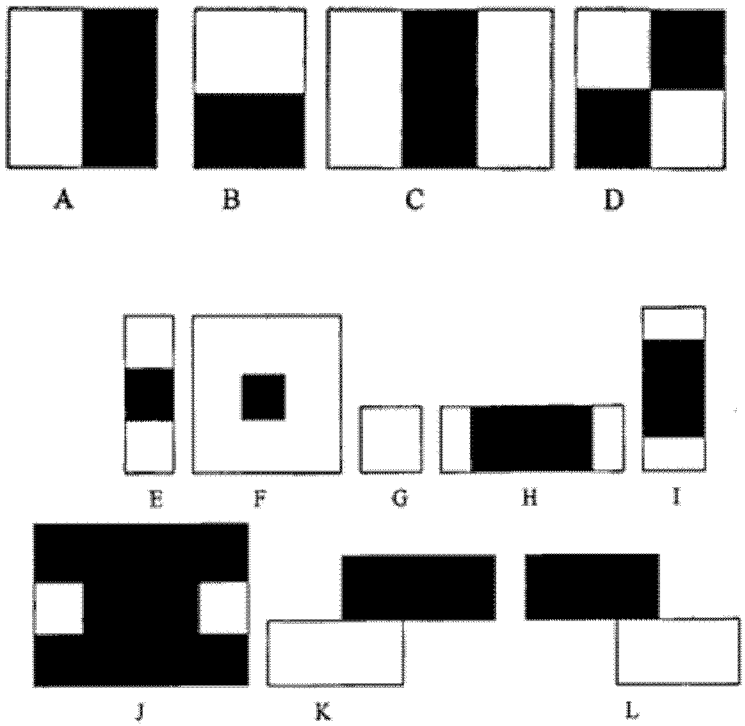

A
B
C
D
E
F
G
H
I
J
K
L

图 4B

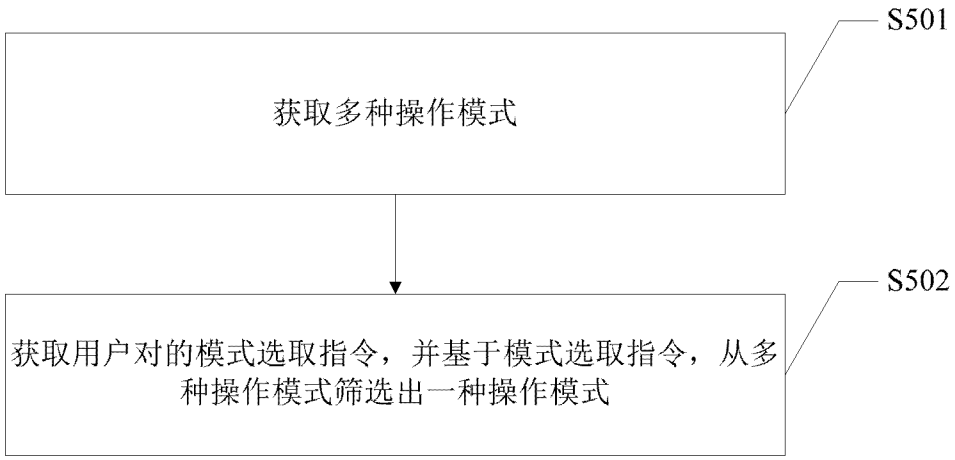

S501
获取多种操作模式
S502
获取用户对的模式选取指令，并基于模式选取指令，从多种操作模式筛选出一种操作模式

图 5

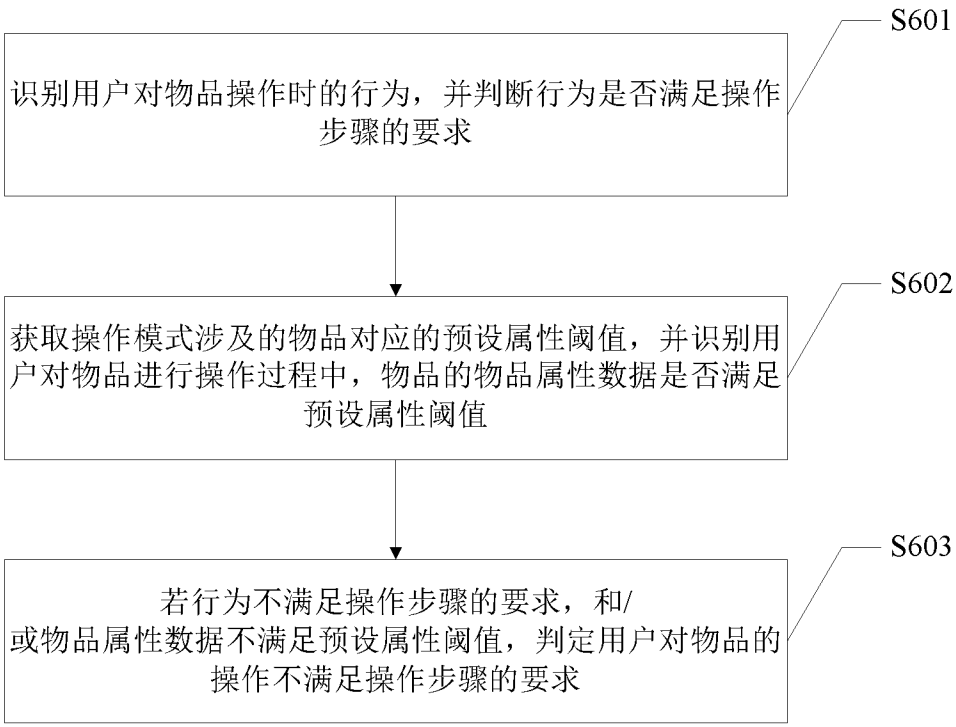

S601
识别用户对物品操作时的行为，并判断行为是否满足操作步骤的要求
S602
获取操作模式涉及的物品对应的预设属性阈值，并识别用户对物品进行操作过程中，物品的物品属性数据是否满足预设属性阈值
S603
若行为不满足操作步骤的要求，和/或物品属性数据不满足预设属性阈值，判定用户对物品的操作不满足操作步骤的要求

图 6

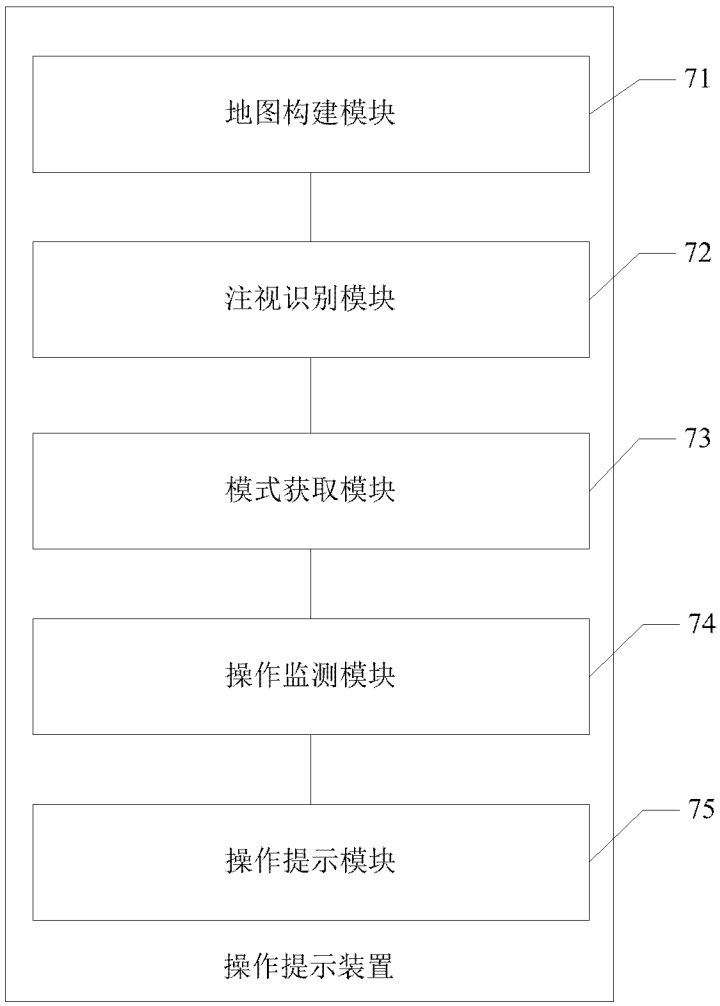

地图构建模块
71
注视识别模块
72
模式获取模块
73
操作监测模块
74
操作提示模块
75
操作提示装置

图 7

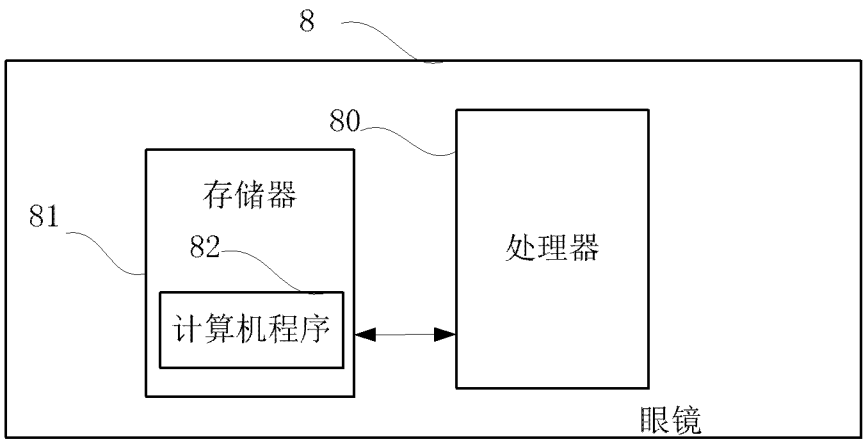

8
80
存储器
81
82
计算机程序
处理器
眼镜

图 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/124352

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/01(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; G06K; G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, WPI, EPODOC, IEEE: 眼镜, 可戴, 可穿戴, 注视, 眼动, 操作, 报警, 警报, 错误, 失误, 指示, 提示, 地图, 彩色, 深度, 图像, 位置, 姿态, 场景, 增强现实, 虚拟现实, 三维, goggles, spectacles, glasses, wearable, eye, gaze, handl+, manipulat+, operat+, alarm, error, indication, denote, description, map, colour, color, depth, image, picture, gesture, scene, 3D, AR, augment, reality

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109782902 A (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 21 May 2019 (2019-05-21) claims 1-10	1-10
Y	CN 106557168 A (SHANGHAI QINGGAN INTELLIGENT TECHNOLOGY CO., LTD.) 05 April 2017 (2017-04-05) description, paragraphs 0033-0052, and figure 1	1-10
Y	CN 104965314 A (CHEN, Lixiao) 07 October 2015 (2015-10-07) description, paragraph 0009	1-10
A	CN 108351522 A (TOBII AB) 31 July 2018 (2018-07-31) entire document	1-10
A	US 2014266988 A1 (EYECAM, LLC) 18 September 2014 (2014-09-18) entire document	1-10

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 February 2020	**25 February 2020**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration** **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/124352

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109782902	A	21 May 2019	None			
CN	106557168	A	05 April 2017	None			
CN	104965314	A	07 October 2015	None			
CN	108351522	A	31 July 2018	US	2017038835	A1	09 February 2017
				EP	3332285	A1	13 June 2018
				WO	2017027352	A1	16 February 2017
				US	2019011986	A1	10 January 2019
				US	2018088668	A1	29 March 2018
US	2014266988	A1	18 September 2014	WO	2014145166	A2	18 September 2014
				US	2016299569	A1	13 October 2016

国际检索报告

国际申请号
PCT/CN2019/124352

A. 主题的分类

G06F 3/01(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F; G06K; G06T

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNKI, CNPAT, WPI, EPODOC, IEEE:眼镜，可戴，可穿戴，注视，眼动，操作，报警，警报，错误，失误，指示，提示，地图，彩色，深度，图像，位置，姿态，场景，增强现实，虚拟现实，三维，goggles, spectacles, glasses, wearable, eye, gaze, handl+, manipulat+, operat+, alarm, error, indication, denote, description, map, colour, color, depth, image, picture, gesture, scene, 3D, AR, augment, reality

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 109782902 A (中国科学院深圳先进技术研究院) 2019年 5月 21日 (2019-05-21) 权利要求1-10	1-10
Y	CN 106557168 A (上海擎感智能科技有限公司) 2017年 4月 5日 (2017-04-05) 说明书第0033-0052段、附图1	1-10
Y	CN 104965314 A (陈丽晓) 2015年 10月 7日 (2015-10-07) 说明书第0009段	1-10
A	CN 108351522 A (托比股份公司) 2018年 7月 31日 (2018-07-31) 全文	1-10
A	US 2014266988 A1 (EYECAM, LLC.) 2014年 9月 18日 (2014-09-18) 全文	1-10

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2020年 2月 16日	2020年 2月 25日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 张雯 电话号码 86-(10)-53961314

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/124352

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	109782902	A	2019年 5月 21日	无			
CN	106557168	A	2017年 4月 5日	无			
CN	104965314	A	2015年 10月 7日	无			
CN	108351522	A	2018年 7月 31日	US	2017038835	A1	2017年 2月 9日
				EP	3332285	A1	2018年 6月 13日
				WO	2017027352	A1	2017年 2月 16日
				US	2019011986	A1	2019年 1月 10日
				US	2018088668	A1	2018年 3月 29日
US	2014266988	A1	2014年 9月 18日	WO	2014145166	A2	2014年 9月 18日
				US	2016299569	A1	2016年 10月 13日