(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
2019 年 2 月 14 日 (14.02.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/029100 A1

(51) 国际专利分类号:
G06F 3/01 (2006.01)

(21) 国际申请号: PCT/CN2017/118923

(22) 国际申请日: 2017 年 12 月 27 日 (27.12.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710668415.X 2017年8月8日 (08.08.2017) CN

(71) 申请人: 山东科技大学 (SHANDONG UNIVERSITY OF SCIENCE AND TECHNOLOGY) [CN/CN]; 中国山东省青岛市经济技术开发区前湾港路579号, Shandong 266590 (CN)。

(72) 发明人: 彭延军 (PENG, Yanjun); 中国山东省青岛市经济技术开发区前湾港路579号, Shandong 266590 (CN)。 王美玲 (WANG, Meiling); 中国山东省青岛市经济技术开发区前湾港路579号, Shandong 266590 (CN)。 王元红 (WANG, Yuanhong); 中国山东省青岛市经济技术开发区前湾港路579号, Shandong 266590 (CN)。 卢新明 (LU, Xinming); 中国山东省青岛市经济技术开发区前湾港路579号, Shandong 266590 (CN)。

(74) 代理人: 青岛智地领创专利代理有限公司 (QINGDAO ZHIDILINGCHUANG PATENT AGENCY CO., LTD); 中国山东省青岛市经济技术开发区香江路110号高科技创业中心215室（联系人：韩玉昆）, Shandong 266555 (CN)。

(54) **Title:** MULTI-INTERACTION IMPLEMENTATION METHOD FOR MINING OPERATION BASED ON VIRTUAL REALITY AND AUGMENTED REALITY

(54) **发明名称:** 一种基于虚拟现实与增强现实的采矿操作多交互实现方法

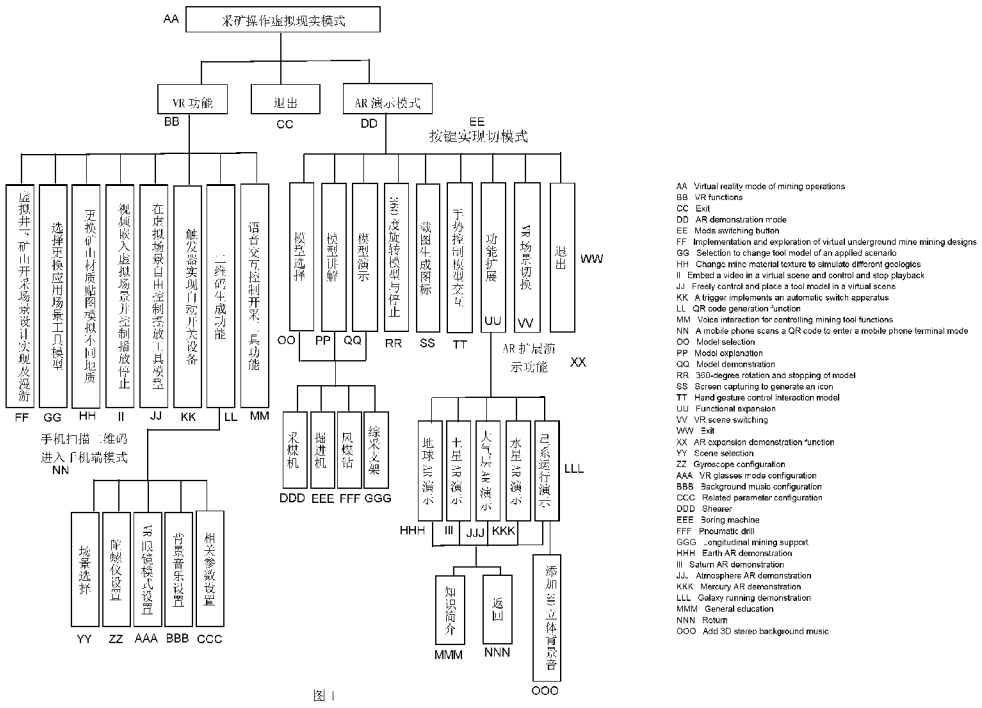

图 1

(57) **Abstract:** The present invention discloses a multi-interaction implementation method for a mining operation based on virtual reality and augmented reality, belonging to the technical field of virtual reality and augmented reality. The method comprises two modes: a virtual reality mode and an augmented reality mode. The following operations can be implemented in a virtual reality scene: displaying a model, selecting and replacing materials, scene exploration, random movement and placement of the model, embedding videos, QR code generation, realizing natural interactions and voice interactions by means of a trigger; The following operations can be

[见续页]

WO 2019/029100 A1

(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, 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))。

implemented in an augmented reality scene: model selection, voice playback, demonstration of model operation dynamics and model rotation and stop control, screen capturing and functional expansions. Various manners of interaction, such as voice control, hand gesture control and control via a keyboard and a mouse cursor can be implemented in the two modes. The method can be applied to virtual simulations of applicable mine operation scenarios. The method is used to train miners in a mining area and students specializing in mining engineering, to reduce training costs and to improve workers' skills, thereby providing a fast and effective means for guiding production, construction, and technical scientific research.

(57) 摘要：一种基于虚拟现实与增强现实的采矿操作多交互实现方法，属于虚拟现实、增强现实技术领域，包括虚拟现实与增强现实两种模式，虚拟现实场景下，可以实现虚拟场景中的模型、材质的选择、更换，场景漫游，任意移动摆放模型，视频嵌入，生成二维码，触发器实现自然交互，语音交互等；增强现实场景下，可以选择模型、播放语音、演示模型运行动态以及控制模型旋转停止、截图和功能扩展；在两种模式下，实现语音控制、手势控制以及键盘鼠标控制的多种交互方式。应用于采矿操作的虚拟仿真应用场景，可用于培训矿区开采工人及采矿工程专业的学生，减少培训资本，提高工人的技能，为指导生产施工和科学技术研究提供了先进、快捷的手段。

一种基于虚拟现实与增强现实的采矿操作多交互实现方法

技术领域

本发明属于虚拟现实、增强现实技术领域，具体涉及一种基于虚拟现实与增强现实的采矿操作多交互实现方法。

背景技术

2016 年被业界称为“虚拟现实元年”，可能会有人误以为这项技术是近几年发展起来的新技术。其实不然，虚拟现实（Virtual Reality，简称 VR）技术兴起于 20 世纪 90 年代，2000 年以后，虚拟现实技术在整合发展中引入了 XML、JAVA 等先进技术，应用强大的 3D 计算能力和交互式技术，提高渲染质量和传输速度，进入了崭新的发展时代。虚拟现实技术是经济和社会生产力发展的产物，有着广阔的应用前景。我国虚拟现实技术的研究起步于 20 世纪 90 年代初。随着计算机图形学、计算机系统工程等的高速发展，虚拟现实技术得到相当的重视。国家广告研究院等多家机构联合发布的《2016 上半年中国 VR 用户行为研究报告》显示，2016 年上半年国内虚拟现实潜在用户达 4.5 亿，浅度用户约为 2700 万，重度用户约 237 万，预计国内虚拟现实市场将迎来爆发式增长。而增强现实（Augmented Reality，简称 AR）技术是在虚拟现实的基础上发展起来的一种新兴技术。其应用领域也非常广泛，其在工业、医疗、军事、市政、电视、游戏、展览等领域都表现出了良好的应用前景。

目前，VR 与 AR 技术不断发展，应用范围也越来越广泛，但是这两种技术更多的应用于军事、娱乐等领域，对于教育、工业、工程等领域的应用，由于领域本身涉及多种物理、地理等多学科因素，还需要更多的研究与发展。对于矿山开采工业领域，我国矿山的地质条件较为复杂，且多为井下开采，在矿山开采过程中，由于开采环境位于地下，工艺工序又颇为复杂，瓦斯、水害等灾害事故时有发生。与此同时，矿山开采又是一个工期长、投资大、安全隐患高的行业，很容易发生安全事故，所以采矿员工的安全培训一直是采矿工作的重中之重。但是，目前存在的传统培训教学系统，基本是理论介绍加以模具展示或者二维图像展示，以课堂讲解为主，辅以简单的动画和音、视频的介绍，实践不足、缺乏真实场景。即便是观看模具也不能很好的掌握工具的实际操作流程。随着技术的不断发展，各种应用于煤矿开采的训练系统也相应开发，但是也存在系统场景真实性差、沉浸性效果不好以及交互性功能少，只能简单的演示等问题。

发明内容

针对现有技术中存在的上述技术问题，本发明提出了一种基于虚拟现实与增强现实的采矿操作多交互实现方法，设计合理，克服了现有技术的不足，具有良好的效果。

为了实现上述目的，本发明采用如下技术方案：

一种基于虚拟现实与增强现实的采矿操作多交互实现方法，采用井下采矿操作多交互仿真系统，该系统包含虚拟现实模式和增强现实模式两种模式；虚拟现实模式包含特定场景的建模、漫游、模型及其材质的更换、视频嵌入虚拟场景、模型移动、应用场景意向交互、二维码生成以及语音交互；增强现实模式包含模型选择、模型讲解、动态模型演示、手势控制模型交互、截图生成图标、360 度旋转以及停止、功能模式切换以及功能扩展；系统有设计两种隐藏菜单，即虚拟现实模式下的更换工具、材质的选择菜单以及增强现实模式下的模型选择类菜单；第一种用户进入特定区域菜单才会显示，离开即会隐藏；第二种点击便可在某处显示二级菜单，再次点击菜单隐藏；

所述的采矿操作多交互实现方法，具体包括如下步骤：

步骤 1：矿山开采操作的整个环境场景的搭建

根据井下采矿操作的真实环境，利用建模工具 3DMax 进行 1：1 等比建模，实现整个井下采矿操作的环境模拟；利用 UE4 引擎对模型进行包括创建、编辑贴图以及材质在内的编辑，

添加物理碰撞，对整体环境进行灯光、效果光照以及特效添加，并进行烘焙、渲染；

步骤 2：虚拟现实应用场景的漫游

在 UE4 引擎中，设置键盘上、下、左、右键，绑定 Up、Down、Right、Left 方向控制函数，为鼠标绑定 Turnaround 控制函数，实现整个井下采矿操作的虚拟现实场景的漫游；

步骤 3：更换井下采矿操作的工具模型以及矿山地质的模拟材质

在虚拟井下矿山开采场景中添加隐藏菜单，当漫游至矿山开采处，会自动出现模型或者材质选择菜单，用户可以根据需求从菜单中选择模型或者材质进行更换；

步骤 4：将视频素材嵌入三维应用场景并控制播放、停止

将视频素材嵌入虚拟现实场景，在三维空间中播放，模拟矿山开采环境的监控显示设备，设置键盘 X 键，绑定 UE4 平台的 MediaPlayer 媒体类，通过 OpenSource 和 Close 函数控制视频的播放和停止；

步骤 5：选择模型并移动到任意位置

通过鼠标选中模型并将模型移动到任意需要进行模拟操作的位置，达到真实场景中的机械移动模拟；

步骤 6：实现应用场景的意向交互

当用户在虚拟现实应用场景中漫游至某一特定位置，系统检测到用户有意向进入，就自动开启环境灯，实现虚拟场景中的自然交互；

步骤 7：二维码生成

绑定键盘的 F 键，添加二维码生成函数，设置键盘按键控制生成二维码功能，用户按键盘 F 键，系统生成含设置好采样点的虚拟场景全景图的二维码；

步骤 8：实现语音交互

用户通过包括正转、反转、升臂、降臂、停止在内的关键词控制虚拟现实场景中的采煤机，模拟其运作效果；

步骤 9：AR 动态演示功能模式切换

用户点击系统右上角的 AR 模式按键切换到 AR 演示模式。

优选地，在步骤 3 中，将模型实例化为具体的 Actor，添加 SetMesh 函数以及 SetMaterial 函数来更换模型和模型材质，设置 Widget Blueprint 用户界面以及 Box collision 碰撞检测，实现三维空间的隐藏菜单功能。

优选地，在步骤 5 中，为要操作的模型添加鼠标事件，通过 GetHitResult 函数将模型选中，然后根据鼠标在三维空间的坐标，改变模型的 SetActorLocation 函数的坐标值，当鼠标再次点击，将此时鼠标 x、y、z 三个方向的坐标值赋给模型，此时 GetHitResult 函数将模型设置为取消选中模式。

优选地，在步骤 6 中，设置 TriggerBox 触发器，当第一人称角色触发 TriggerBox，系统检测到用户有意向进入某区域，便会自动启用此区域的某个设备。

优选地，在步骤 7 中，用户按键盘 F 键，系统生成含设置好采样点的虚拟场景全景图二维码，用户用手机扫描二维码，跳转到手机端的虚拟应用场景展示页面，在手机端，用户可以启用陀螺仪，切换到 VR 分屏模式，设置好手机参数，便可用 VR 眼镜体验虚拟井下采矿操作环境场景，实现 720 度的视角展示，还能够实现手机端的多场景、多角度的漫游体验。

优选地，在步骤 8 中，语音识别基于 Pocket-sphinx 库实现，通过改进中文关键字字典，经过预处理、特征提取、声学模型训练、语言模型训练以及语音解码和搜索实现识别功能，最后经过 UE4 引擎的编写功能控制函数实现三维空间里语音对模型的控制；语音识别的具体实现步骤如下：

步骤 8.1：预处理

对输入的原始语音信号进行处理，滤除掉其中的不重要的信息以及背景噪声，并对语音信号的端点检测、语音分帧和预加重进行处理；

通过一阶 FIR 高通数字滤波器来实现预加重，一阶 FIR 高通数字滤波器的传递函数为：

$$\mathbf{H}(\boldsymbol{z}) = \mathbf{1} - \boldsymbol{a}\boldsymbol{z}^{-\mathbf{1}};$$

其中，a 为预加重滤波器的系数，取值范围为 0.9~1.0，若设 n 时刻的语音采样值为 x(n)，则预加重后的信号为

$$\mathrm{y(n)} = \mathrm{x(n)} - \mathrm{a} * \mathrm{x(n-1)};$$

步骤 8.2：特征提取

通过梅尔频率倒谱系数（MFCC）的方法来进行特征提取；具体按照如下步骤进行：

步骤 8.2.1：利用人听觉的临界带效应，采用 MEL 倒谱分析技术对语音信号处理得到 MEL 倒谱系数矢量序列；

步骤 8.2.2：用 MEL 倒谱系数矢量序列表示输入语音的频谱，在语音频谱范围内设置若干个具有三角形或正弦形滤波特性的带通滤波器；

步骤 8.2.3：通过带通滤波器组，求各个带通滤波器的输出数据；

步骤 8.2.4：对各个带通滤波器的输出数据取对数，并做离散余弦变换(DCT)；

步骤 8.2.5：得到 MFCC 系数；求解公式如下：

$$\boldsymbol{C_i} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{P}} \mathbf{log}\boldsymbol{F(k)}\,\mathbf{cos}[\boldsymbol{\pi(k-0.5)i/P}] \qquad \boldsymbol{i} = \mathbf{1,2,}\ldots\boldsymbol{P};$$

其中，$\boldsymbol{C_i}$为特征参数，$\boldsymbol{k}$为三角滤波器的个数，F(k)为各个滤波器的输出数据，P为滤波器阶数，$\boldsymbol{i}$为数据长度；

步骤 8.3：声学模型训练

根据训练语音库的特征参数训练出声学模型参数；

在识别时可以将待识别的语音的特征参数同声学模型进行匹配，得到识别结果；本文采用混合高斯模型-隐马尔科夫模型（GMM-HMM）作为声学模型，具体包括如下步骤：

步骤 8.3.1：求出混合高斯模型的联合概率密度函数：

$$\mathrm{P(x)} = \sum_{m=1}^{M} \frac{C_m}{(2\pi)^{D/2}|\sum_m|^{1/2}} exp\left[-\frac{1}{2}(x-u_m)^T \sum\nolimits_m^{-1}(x-u_m)\right]$$

$$= \sum_{\boldsymbol{m=1}}^{\boldsymbol{M}} C_m\, \boldsymbol{N}(\boldsymbol{x};\boldsymbol{u}_m,\sum\boldsymbol{m})$$

其中，M表示混合高斯模型中高斯的个数，C_m表示权重，$\boldsymbol{u}_m$表示均值，$\sum\boldsymbol{m}$表示协方差矩阵，D 为观测矢量维数；利用最大期望值算法（EM）对混合高斯模型参数变量

$\boldsymbol{\Theta} = \{C_m,\ \boldsymbol{u}_m,\ \sum\boldsymbol{m}\}$进行估计，利用如下公式求解：

$$\boldsymbol{c_m^{(j+1)}} = \frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{t=1}}^{\boldsymbol{N}} \boldsymbol{h_m^{(j)}(t)}$$

$$\boldsymbol{u_m^{(j+1)}} = \frac{\sum_{t=1}^{N} \boldsymbol{h_m^{(j)}(t)x^{(t)}}}{\sum_{t=1}^{N} \boldsymbol{h_m^{(j)}(t)}}$$

$$\sum\nolimits_{\boldsymbol{m}}^{(j+1)} = \frac{\sum_{t=1}^{N} \boldsymbol{h_m^{(j)}}\left[\boldsymbol{x^{(t)}} - \boldsymbol{u_m^{(j)}}\right]\left[\boldsymbol{x^{(t)}} - \boldsymbol{u_m^{(j)}}\right]^T}{\sum_{t=1}^{N} \boldsymbol{h_m^{(j)}(t)}}$$

其中，$\boldsymbol{j}$是当前迭代轮数，$\boldsymbol{N}$表示训练数据集中元素的个数，$\boldsymbol{x^{(t)}}$为$\boldsymbol{t}$时刻的特征向量，$\boldsymbol{h_m(t)}$表示 $\boldsymbol{t}$时刻C_m的后验概率；GMM 参数通过 EM 算法进行估计，可以使其在训练数据上生成语音观察特征的概率最大化；

步骤 8.3.2：求解 HMM 的三个主要组成部分

设状态序列为$q_1, q_2, \ldots, q_N$，令转移概率矩阵$\mathbf{A} = [a_{ij}]\ i,j \in [1,N]$，则求出的马尔科夫链状态间的跳转概率为：$a_{ij} = P(q_t = j | q_{t-1} = i)$；马尔科夫链的初始概率$\boldsymbol{\pi} = [\pi_i]\ i \in [1,N]$，其中，$\pi_i = P(q_1 = i)$；令每个状态的观察概率分布$b_i(o_t) = P(o_t | q_t = i)$，采用 GMM 模型来描述状态的观察概率分布；根据步骤 8.3.1，求解公式为：

$$b_i(o_t) = \sum_{m=1}^{M} \frac{C_{i,m}}{(2\pi)^{D/2}|\sum_{i,m}|^{1/2}} exp\left[-\frac{1}{2}(o_t - u_{i,m})^T \sum_{i,m}^{-1}(o_t - u_{i,m})\right]$$

其中，N为状态个数，i、j表示状态，a_{ij}表示$t-1$时刻从i状态跳转到t时刻j状态的概率，o_t为t时刻的观测值，$C_{i,m}$为混合系数，表示不同高斯之间的权重，$u_{i,m}$表示不同高斯之间的均值，$\sum_{i,m}$表示不同高斯之间的协方差矩阵；HMM 的参数通过 Baum-Welch 算法进行估计得出，最后生成声学模型文件；

步骤 8.4：语言模型训练

采用 N-Gram 模型实现语言模型的训练；在一个语句中第i个词出现的概率，条件依赖于它前面的 *N*-1 个词，即将一个词的上下文定义为该词前面出现的 *N*-1 个词，其表达式为：

$$\mathrm{P}(w_i = \mathrm{w}|\mathrm{c}) = \mathrm{P}(w_i = \mathrm{w}|w_{i-N+1}^{i-1})$$

使用条件概率公式 S 将上述表达式替换成如下公式：

$$\mathrm{P}(\mathrm{sentence}) = P(w_1) * P(w_2|w_1) * P(w_3|w_2) \ldots * P(w_n|w_1, w_2, \ldots, w_{n-1})$$

其中，$\mathrm{P}(w_1)$是w_1在文章中出现的概率，$\mathrm{P}(w_1, w_2)$是w_1, w_2连续出现的概率，$\mathrm{P}(w_2|w_1)$是已知w_1已出现的情况下出现w_2的概率，假设识别 sentence 的概率用$\mathrm{P}(s)$表示，$\mathrm{P}(s) = P(w_1, w_2, \ldots, w_n)$ 表示单词集$w_1, w_2, \ldots, w_n$连续出现并生成 S 的概率；

通过马尔科夫假设精简成如下公式：

$$\mathrm{P}(\mathrm{sentence}) = P(w_1) * P(w_2|w_1) * P(w_3|w_2) \ldots * P(w_n|w_{n-1})$$

其中，$\mathrm{P}(w_i|w_{i-1}) = \mathrm{P}(w_{i-1}, w_i)/\mathrm{P}(w_i)$，$\mathrm{P}(w_{i-1}, w_i)$和$\mathrm{P}(w_i)$都可以从语料统计出来，最终就能得到$\mathrm{P}(\mathrm{sentence})$，语言模型存储$\mathrm{P}(w_{i-1}, w_i)$的概率统计值，通过求出 P(sentence)的最大值来实现整个识别过程；

步骤 8.5：语音解码和搜索算法

针对输入的语音信号，根据已经训练好的声学模型、语言模型及利用 g2p 工具创建好的字典映射文件建立一个识别网络，根据搜索算法在该网络中寻找最佳的一条路径，这个路径就是能够以最大概率输出该语音信号的词串，这样就确定这个语音样本所包含的文字，本文采用 Viterbi 算法实现语音解码，具体过程如下：

步骤 8.5.1：输入 HMM 模型的参数和观测序列$\mathrm{O} = \{\mathrm{o}_1, \mathrm{o}_2, \ldots, \mathrm{o}_\mathrm{T}\}$，则 t=1 时所有的状态概率为：

$$\delta_1(\mathrm{i}) = \pi_\mathrm{i} \mathrm{b}_\mathrm{i}(\mathrm{o}_1)$$
$$\psi_1(\mathrm{i}) = 0$$

步骤 8.5.2：逐渐递推到 *t*=2,3,…,T，则为：

$$\delta_t(i) = \max_{1 \le j \le N}\left[\delta_{t-1} a_{ij}\right] b_i(o_t)$$

$$\psi_t(i) = \mathrm{argmax}_{1 \le j \le N}\left[\delta_{t-1}(j) a_{ji}\right]$$

步骤 8.5.3：终止遍历：

$$P^* = \max_{1\le i\le N} \delta_T(i)$$

$$i_T^* = \underset{1\le i\le N}{\operatorname{argmax}}[\delta_T(i)]$$

步骤 8.5.4：回溯最优路径，t=T-1,T-2,…,1；

$$i_t^* = \psi_{t+1}(i_{t+1}^*)$$

步骤 8.5.5：输出最优隐状态路径$I^* = (i_1^*, i_2^*, \dots, i_T^*)$；

其中，$\delta_t(i)$是递推到 t 时刻，最优路径经过的所有结点的联合概率，$\psi_t(i)$是 t 时刻的隐状态，T为时间，P^*为最优路径的概率，i_T^*为最优路径的终结点。

优选地，a 取 0.97。

优选地，在步骤 9 中，具体包括如下步骤：

步骤 9.1：模型选择

对采煤机模型、掘进机模型、风煤钻模型以及综采支架模型进行选择，每一类模型都是对真实采煤工具的 1:1 建模模拟；

步骤 9.2：模型讲解

用户通过选择模型后，再通过此菜单选择需要学习的工具模型选项，系统会播放对应的语音讲解，再次点击按键语音停止；

步骤 9.3：模型演示

将在 3DMax 建模过程中制作的工具模拟运行动画导入到 Unreal Engine 引擎中，设置相应的选择菜单，点击便可在 AR 模式下演示相应采煤工具的运行状态；

步骤 9.4：截图生成图标

在 AR 模式的主菜单，添加一个按钮，绑定摄像机的截图功能，在菜单右侧添加滚动菜单栏，当截图函数成功触发，截图通过设置好的动态材质转换函数，显示到右侧滚动菜单栏，演示过程中，用户点击截图按键，系统会在界面一侧生成图标；

步骤 9.5：旋转

将设置的模型实例化为一个 Actor，添加 Rotation 函数，实现模型顺时针旋转；

步骤 9.6：功能扩展

添加二级 UI，控制 Map 切换，实现包括地球、土星、水星、含大气层星球以及星系在内的运行演示功能；添加 WidgetBlueprint 编码实现了知识简介面板的显示或隐藏；设计返回键可以回到 AR 编辑主模块；

步骤 9.7：动态手势控制模型，真实环境与虚拟模型叠加，手势与模型进行交互控制，具体包括如下步骤：

步骤 9.7.1：初始化视频捕捉，读取标志文件和摄像相机参数；

步骤 9.7.2：抓取视频帧图像；

步骤 9.7.3：执行探测标记以及识别视频帧中的标记模板，并利用 OpenCV 库函数对获取的视频帧图像进行运动检测，判断是否检测到运动轨迹；

若：判断结果是检测到运动轨迹，则执行步骤 9.7.4：

或判断结果是没有检测到运动轨迹，则继续执行探测标记以及识别视频帧中的标记模板，然后执行步骤 9.7.12；

基于颜色直方图与背景差分进行运动检测，对采集的帧以及对每帧运动检测后得到除运动手势区域外的像素做背景更新，公式如下；

$$u_{t+1} = \begin{cases} u, & I_f(x,y) = 1 \\ au_t + (1-a)I_t, & \text{其他} \end{cases}$$

其中，u_t为背景图像相应的像素点，u_{t+1}为更新后的背景图像像素点；I_t为当前帧图像的

像素点，I_f是当前帧图像像素点的掩码值，即是否做背景更新；$a \in [0,1]$为背景图像模型更新速度；

步骤 9.7.4：对图像进行包括去噪在内的预处理；

通过运动检测步骤，如果检测到有运动信息，则开始对含有运动手势的视频帧图像进行预处理：通过 OpenCV 的 medianBlur 函数对图像进行中值滤波，去除椒盐噪声；

步骤 9.7.5：转换到 HSV 空间；

通过 cvtColor 函数对图像进行颜色空间转换，得到其 HSV 空间的数据，并对 HSV 空间中的亮度v值重新设定如下式所示：

$$\begin{cases} v_{temp} = \begin{cases} r/g, & g \neq 0 \\ 4, & g = 0 \end{cases} \\ v = \begin{cases} 10r/g, v_{temp} \leq 2 \\ 255, \quad v_{temp} > 2 \end{cases} \end{cases}$$

其中，r、g为肤色区域的红色与绿色像素，且$r > g$；

步骤 9.7.6：分割手区域；

步骤 9.7.7：进行形态学处理，去除杂点；

将得到的运动二值图和通过反投影得到的二值图相与，并进行图像形态学闭操作得到比较完整的运动肤色手势二值图；并去除图像中的杂点；

步骤 9.7.8：获取手轮廓；

通过初步的形态学操作，去除噪声，并使手的边界更加清晰后，通过 OpenCV 的 findContours 函数得到手势轮廓，然后进行去除伪轮廓操作；

步骤 9.7.9：画出手轮廓，标定信息；

步骤 9.7.10：轮廓信息比较，设置方向向量；

将每一帧得到的轮廓进行比较，设定比较条件，通过比较给方向标志变量赋值；

步骤 9.7.11：对模型根据矢量坐标进行受力模拟，实现动态手势与虚拟模型的交互；

动态手势通过轮廓判断后，根据不同的判断结果对虚拟模型进行受力模拟操作，根据轮廓判断过程中方向标记的值，模型在三维空间的坐标值将会进行 x、y、z 三个坐标轴上的相乘计算，通过坐标值的改变，实现模型位置的改变而达到受力的模拟；

步骤 9.7.12：计算摄像头相对于探测到的标记的转换矩阵；

步骤 9.7.13：在探测到的标记上叠加虚拟物体，并返回执行步骤 9.7.2，实现真实环境与虚拟模型的叠加显示。

本发明所带来的有益技术效果：

（1）本发明的三维模型采用等比例建立，材质贴图通过 UE4 引擎平台的编辑贴近真实，应用场景的环境灯光采用真实灯光模拟烘焙渲染。整个虚拟现实场景都更真实，沉浸感极强。

（2）本发明通过技术方案实现了多种功能交互，例如在虚拟井下开采场景漫游的过程中通过隐藏菜单更换工具模型，更换矿山材质来模拟不同的开采地质，自由移动开采工具的位置，以及视频信息嵌入机器显示屏展现真实场景，利用语音功能实现控制采煤机的正转、反转、升臂、降臂、停止等。

（3）本发明还通过生成二维码功能，将 PC 端展示连接到手机端展示，手机端功能更是可以利用手机内置陀螺仪，产生重力感应，如果在设置成 VR 眼镜模式便可利用简单的 VR 眼镜体验实时的场景沉浸感。

（4）本发明还利用 AR 开发 SDK—ARToolKit 实现了 AR 动态演示功能，通过 AR 编辑演示功能，用户可是实时选择采矿工具模型，进行 360 旋转展示、语音讲解以及动态运行展示，截图保存等，更重要的是其将工具模型以 AR 的模式展示，虚拟模型与真实环境结合的展示效果，这不仅能展现模型的直观立体性，更能展现其真实性，使其具有更好的学习、教

育的效果。

（5）本发明的 AR 模块，除了其动态演示功能，更是对视频流添加了处理，当动态手势进入摄像头视角，它会产生与模型的交互，手从远到近的动态会传递给模型一个三维空间向前的一个模拟力，从上到下的动态会给模型一个向上的一个模拟力，向前翻转手的动态会给模型一个向下的模拟了，同样，如果手扭动或者左右倾斜，便会给模型一个具有矢量方向的模拟力。

（6）本发明在 AR 模块除了在煤矿应用场景的功能实现，还扩展了 AR 在天文学领域的展示功能。添加地球、土星、水星、含有动态大气层的星球以及星系的 AR 展示功能，与此同时将知识简介面板显示功能添加到此 AR 展示的模块中，丰富了 AR 在教育展示领域的应用。

附图说明

图 1 是本发明实现的整体功能结构图。

图 2 是本发明生成二维码功能的示意图。

图 3 是本发明语音识别实现交互功能的原理图。

图 4 是本发明 AR 模式实现的原理图。

图 5 是本发明动态手势交互功能实现的流程图。

具体实施方式

下面结合附图以及具体实施方式对本发明作进一步详细说明：

本发明提供一种基于虚拟现实与增强现实的采矿操作多交互实现方法。结合附图 1 可以了解本发明所包含的整个技术功能。其具体实施步骤如下：

步骤 1：井下矿山开采作业的整个环境场景搭建。利用 3DMax 建模工具根据真实采矿操作环境创建相关模型。将模型分类导入 UE4 引擎，通过 UE4 平台，对模型进行材质编写、模拟自然灯光、环境光，添加物理碰撞检测，对系统进行参数调整，烘焙渲染。

步骤 2：在虚拟应用场景添加第一人称角色，给角色添加鼠标键盘控制事件。将键盘的上下左右键绑定 Up、Down、Right、Left 函数，控制第一人称角色在虚拟三维空间的坐标改变，实现漫游。给鼠标添加 Turnaround 函数，控制第一人称视角在虚拟三维空间的 720 度旋转。

步骤 3：设置交互菜单，实现更换井下采矿操作的工具模型、矿山地质材质等功能交互。首先创建一个 Widget Blueprint 用户界面，设置菜单选项，为选项添加点击事件。然后给模型添加 Box collision 碰撞检测区域，当角色进入 Box collision 碰撞检测区域。创建的 Widget Blueprint 用户界面显示。离开 Box collision 碰撞检测区域，Widget Blueprint 用户界面隐藏。将采煤机模型实例化为一个 Actor，并添加 SetMesh 函数，实现更换其他工具模型。同理，将三维空间里的矿山地质模型添加 SetMaterial 函数，实现更换材质。本发明设置四类开采工具模型供用户选择，以及将矿山地质设置成材质可选择模式，通过显示的样式菜单更换模型、材质。更换完毕，离开检测区域，菜单自动隐藏，不影响整体漫游视觉效果，又能达到实时交互的功能。

步骤 4：视频嵌入，在三维空间播放，模拟矿山开采环境的监控显示设备。本发明设置键盘 X 键绑定 UE4 平台的 MediaPlayer 媒体类，通过 Open-Source 和 Close 函数实现控制视频流的播放和停止。此操作可以模拟井下矿山开采控制设备的屏幕显示，以及实时环境监控的画面显示，凸显三维场景的真实性与动态性，使模拟的虚拟场景更加贴近现实。

步骤 5：选择模型可以拖动到任意用户想要放置的位置，以及实现设备自动开启的意向交互功能。为要操作的模型添加鼠标事件，通过 GetHitResult 函数将模型选中，然后根据鼠标在三维空间的坐标，改变模型的 SetActorLocation 函数的坐标值。当鼠标再次点击，将此时鼠标 x、y、z 三个方向的坐标值赋给模型，此时 GetHitResult 函数将模型设置为取消选中模式。本实施例用户可以点击场景中的采煤机模型，将其放到采矿操作场景的其他开采位置。

系统在特定区域添加 TriggerBox 触发器，第一人称角色进入此区域，触发 TriggerBox 触

发器，相应下一个区域的环境灯控制函数 SetVisible 触发，灯被打开，从而实现了本发明设置的自动感应灯功能。这也是本发明设计的检测人意向的功能，从而实现更自然的系统交互。

步骤 6：二维码生成功能。单一的 PC 端展示不能满足多用户的体验，本发明通过添加二维码生成，扫描二维码便可实现多用户手机端的展示，通过二维码连接，手机跳转到煤矿开采作业的全景展示页面。在手机端，用户可以启用陀螺仪，切换到 VR 分屏模式，设置好手机参数，便可用 VR 眼镜体验虚拟井下煤矿开采环境，实现 720 度的视角展示。与此同时，可以实现手机端的多场景、多角度的漫游体验。本功能主要是通过绑定键盘的 F、V 键，添加二维码生成与隐藏函数。在 UE4 引擎中添加场景 6 个 Point 采集点，通过采集点位置生成全景图，再将信息和相关手机端设置以生成二维码形式生成网络连接，实现端与端的转换。此功能实现的流程如图 2 所示。

步骤 7：实现语音控制功能。本发明利用 Pocket-sphinx 实现中文的关键字识别。具体的语音控制实现原理流程如图 3 所示，本发明在以采煤机模型创建的 Actor 上添加语音识别函数，通过在系统初始化后启用语音识别类，并保存对此类的引用。之后创建并绑定一个方法到语音识别函数 OnWordSpoken，每当用户说出设置好的控制词语时，便会触发此方法，通过关键词匹配实现采煤机的正转、反转、升臂、降臂以及停止等相关控制。本方法实现的语音识别是基于美国卡内基梅隆大学开发的英语语音识别系统 Sphinx 改进而实现的。本发明的语音识别方法是大量词汇、非特定人、连续中文音节的孤立词识别方法。能够很好的识别不同人发出的设定词汇。最终通过 UE4 的编码技术，实现了语音词汇识别后与匹配词对应动作控制函数的触发，实现模型的相应动作控制。此识别体系包括语音预处理、特征提取、声学模型训练、语言模型训练和语音解码五个部分。以下是语音识别的具体流程：

步骤 7.1：预处理。

对输入的原始语音信号进行处理，滤除掉其中的不重要的信息以及背景噪声，并进行语音信号的端点检测、语音分帧、预加重等处理。语音信号的预加重，目的是为了对语音的高频部分进行加重，去除口唇辐射的影响，增加语音的高频分辨率。一般通过传递函数为 $\mathbf{H}(\boldsymbol{z}) = \mathbf{1} - \boldsymbol{az^{-1}}$一阶 FIR 高通数字滤波器来实现预加重，$\boldsymbol{a}$为预加重滤波器的系数，取值范围一般在 0.9~1.0，本文取 0.97。设 n 时刻的语音采样值为 x(n)，预加重后的信号为

$$\mathrm{y(n)} = \mathrm{x(n)} - \mathrm{a} * \mathrm{x(n-1)}$$

步骤 7.2：特征提取。

本文使用梅尔频率倒谱系数（MFCC）的方法来提取。MFCC 参数是基于人的听觉特性的，他利用人听觉的临界带效应，采用 MEL 倒谱分析技术对语音信号处理得到 MEL 倒谱系数矢量序列，用 MEL 倒谱系数表示输入语音的频谱。在语音频谱范围内设置若干个具有三角形或正弦形滤波特性的带通滤波器，然后将语音能量谱通过该滤波器组，求各个滤波器输出，对其取对数，并做离散余弦变换(DCT)，即可得到 MFCC 系数。求解公式如下：

$$\boldsymbol{C_i} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{P}} \mathbf{log}\boldsymbol{F(k)}\,\mathbf{cos}[\boldsymbol{\pi(k-0.5)i/P}] \qquad \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots \boldsymbol{P};$$

其中，$\boldsymbol{C_i}$为特征参数，$\boldsymbol{k}$为三角滤波器的个数，F(k)为各个滤波器的输出数据，P为滤波器阶数，$\boldsymbol{i}$为数据长度。

步骤 7.3：声学模型训练。

根据训练语音库的特征参数训练出声学模型参数。在识别时可以将待识别的语音的特征参数同声学模型进行匹配，得到识别结果。本文采用混合高斯模型-隐马尔科夫模型（GMM-HMM）作为声学模型。

步骤 7.3.1：求出混合高斯模型的联合概率密度函数：

$$\mathrm{P(x)} = \sum_{m=1}^{M} \frac{C_m}{(2\pi)^{D/2}|\sum_m|^{1/2}} exp\left[-\frac{1}{2}(x-u_m)^T \sum\nolimits_m^{-1}(x-u_m)\right]$$

$$= \sum_{m=1}^{M} C_m N(x; u_m, \sum_m)$$

其中，M表示混合高斯模型中高斯的个数，C_m表示权重，u_m表示均值，$\sum_m$表示协方差矩阵，D 为观测矢量维数。利用最大期望值算法（EM）对混合高斯模型参数变量：$\Theta = \{C_m, u_m, \sum_m\}$进行估计，利用如下公式求解：

$$c_m^{(j+1)} = \frac{1}{N}\sum_{t=1}^{N} h_m^{(j)}(t)$$

$$u_m^{(j+1)} = \frac{\sum_{t=1}^{N} h_m^{(j)}(t) x^{(t)}}{\sum_{t=1}^{N} h_m^{(j)}(t)}$$

$$\sum_m^{(j+1)} = \frac{\sum_{t=1}^{N} h_m^{(j)}\left[x^{(t)} - u_m^{(j)}\right]\left[x^{(t)} - u_m^{(j)}\right]^T}{\sum_{t=1}^{N} h_m^{(j)}(t)}$$

其中，j是当前迭代轮数，N表示训练数据集中元素的个数，$x^{(t)}$为t时刻的特征向量，$h_m(t)$表示 t时刻C_m的后验概率。GMM 参数通过 EM 算法进行估计，可以使其在训练数据上生成语音观察特征的概率最大化。

步骤 7.3.2：求解 HMM 三个主要组成部分。

设状态序列为$q_1, q_2, \ldots, q_N$，令转移概率矩阵$A = [a_{ij}]\ i, j \in [1, N]$，则求出的马尔科夫链状态间的跳转概率为：$a_{ij} = P(q_t = j | q_{t-1} = i)$；马尔科夫链的初始概率$\pi = [\pi_i]\ i \in [1, N]$，其中，$\pi_i = P(q_1 = i)$；令每个状态的观察概率分布$b_i(o_t) = P(o_t | q_t = i)$，采用 GMM 模型来描述状态的观察概率分布；根据步骤 7.3.1，求解公式为：

$$b_i(o_t) = \sum_{m=1}^{M} \frac{C_{i,m}}{(2\pi)^{D/2} |\sum_{i,m}|^{1/2}} exp\left[-\frac{1}{2}(o_t - u_{i,m})^T \sum_{i,m}^{-1} (o_t - u_{i,m})\right]$$

其中，N为状态个数，i、j表示状态，a_{ij}表示$t-1$时刻从i状态跳转到t时刻j状态的概率，o_t为t时刻的观测值，$C_{i,m}$为混合系数，表示不同高斯之间的权重，$u_{i,m}$表示不同高斯之间的均值，$\sum_{i,m}$表示不同高斯之间的协方差矩阵；HMM 的参数通过 Baum-Welch 算法进行估计得出，最后生成声学模型文件；

步骤 7.4：语言模型训练。

语言模型是用来约束单词搜索，语言建模能够有效的结合汉语语法和语义的知识，描述词之间的内在关系，从而提高识别率，减少搜索范围。本文采用 N-Gram 模型实现语言模型的训练。在一个语句中第i个词出现的概率，条件依赖于它前面的 N-1 个词，即将一个词的上下文定义为该词前面出现的 N-1 个词，其表达公式为：

$$P(w_i = w | c) = P(w_i = w | w_{i-N+1}^{i-1})$$

本文取 N=2 和 N=3，也就是通过前一个或两个单词来判定当前单词出现的概率$P(w_2|w_1)$，$P(w_3|w_2, w_1)$。

简单的说，语言模型就是统计语料得到的模型，语料是用于训练的文本库，字典文件存放的就是训练用的语料和对应发言。语言模型就是表达的语料的组合概率。如设$P(w_1)$是w_1在

文章中出现的概率，$P(w_1, w_2)$是w_1, w_2连续出现是概率，$P(w_2|w_1)$是已知w_1已出现的情况下出现w_2的概率，假设识别 sentence 的概率用$P(s)$表示，$P(s) = P(w_1, w_2, \dots, w_n)$ 表示单词集$w_1, w_2, \dots, w_n$连续出现并生成 S 的概率，使用条件概率公式 S 把整个公式替换成：

$$P(\text{sentence}) = P(w_1) * P(w_2|w_1) * P(w_3|w_2) \dots * P(w_n|w_1, w_2, \dots, w_{n-1})$$

再用马尔科夫假设精简成：

$$P(\text{sentence}) = P(w_1) * P(w_2|w_1) * P(w_3|w_2) \dots * P(w_n|w_{n-1})$$

我们知道，$P(w_i|w_{i-1}) = P(w_{i-1}, w_i)/P(w_i)$，$P(w_{i-1}, w_i)$和$P(w_i)$都可以从语料统计出来，最终就能得到P(sentence)。语言模型存储$P(w_{i-1}, w_i)$的概率统计值，通过求出 P(sentence)的最大值来实现整个识别过程。

步骤 7.5：语音解码和搜索算法。

针对输入的语音信号，根据已经训练好的声学模型、语言模型及字典建立一个识别网络，根据搜索算法在该网络中寻找最佳的一条路径，这个路径就是能够以最大概率输出该语音信号的词串，这样就确定这个语音样本所包含的文字。本文采用 Viterbi 算法实现语音的解码。具体过程如下：

(1)输入 HMM 模型的参数和观测序列$O = \{o_1, o_2, \dots, o_T\}$,则 t=1 时所有的状态概率为：

$$\delta_1(i) = \pi_i b_i(o_1)$$

$$\psi_1(i) = 0$$

(2)逐渐递推到 t=2,3,…,T, 则为：

$$\delta_t(i) = \max_{1 \le j \le N} [\delta_{t-1} a_{ij}] b_i(o_t)$$

$$\psi_t(i) = \operatorname{argmax}_{1 \le j \le N} [\delta_{t-1}(j) a_{ji}]$$

(3)终止遍历：

$$P^* = \max_{1 \le i \le N} \delta_T(i)$$

$$i_T^* = \operatorname{argmax}_{1 \le i \le N} [\delta_T(i)]$$

(4)回溯最优路径，t=T-1,T-2,…,1；

$$i_t^* = \psi_{t+1}(i_{t+1}^*)$$

输出最优隐状态路径$I^* = (i_1^*, i_2^*, \dots, i_T^*)$。其中，$\delta_t(i)$是递推到 t 时刻，最优路径经过的所有结点的联合概率，$\psi_t(i)$是 t 时刻的隐状态，T为时间，P^*为最优路径的概率，i_T^*为最优路径的终结点。最后通过最优路径实现语音识别。

用户说出升臂、降臂、正转、反转以及停止语音后，仿真系统实现采煤机的相应操作，系统识别出用户说出的关键字后会在界面的左上角显示。

步骤 8：AR 动态演示功能模式切换。

在界面设置一个 widget blueprint，添加 openLevel 函数，切换到新的 Map，即 AR 模式。进入 AR 演示模式，此模式具体实现采煤过程中的工具模型演示讲解，从而实现 AR 技术的学习、教育应用功能。

步骤 9：AR 模式下的模型选择、模型讲解以及动态演示。

本发明的 AR 动态演示模块，用户界面为了更简洁和便于 AR 展示，设计二级隐含菜单，本实施例是将模型选择、模型讲解、模型演示以及功能扩展的附加子功能选择设计成隐藏的二级菜单，模型选择分为采煤机、掘进机、风煤钻、综采支架等模型，用户选择完毕，子菜单隐藏即可，模型讲解、模型动态演示以及功能扩展菜单亦是如此实现。具体实现包含内容可参照图 1。本文以 NFT(自然图片追踪,Natural Feature Tracking)为例实现 AR 技术，其原理如图 4 所示，具体流程如下：

步骤 9.1：通过摄像头校准标定，获取到因为摄像头制造工艺偏差而造成的畸变参数，也

就是摄像头内参(intrinsic matrix)，来复原相机模型的 3D 空间到 2D 空间的一一对应关系。

步骤 9.2：根据摄像头本身的硬件参数，我们可以计算出相应的投影矩阵(Projection Matrix)。

步骤 9.3：对待识别的自然图片进行特征提取，获取到一组特征点{p}。

步骤 9.4：实时对摄像头获取到的图像进行特征提取，也是一组特征点{q}。

步骤 9.5： 使用 ICP(Iterative Closest Point)算法来迭代求解这两组特征点的 R、T 矩阵(Rotation&Translation)，即 Pose 矩阵，也就是图形学中常说的模型视图矩阵(Model View Matrix)。假设三维空间的两个点为：$\overrightarrow{p_i} = (x_i, y_i, z_i)$，$\overrightarrow{q_j} = (x_j, y_j, z_j)$，他们的欧氏距离为：

$$\mathrm{d}(\overrightarrow{p_i}, \overrightarrow{q_j}) = \left|\left|\overrightarrow{p_i} - \overrightarrow{q_j}\right|\right| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

为求 p 和 q 变化的矩阵 R 和 T，对于$\overrightarrow{q_j} = R\overrightarrow{p_i} + T$，其中$i, j$=1,2,…,$N$，利用最小二乘法求出最优解。使：

$$\mathrm{E} = \sum_{i=1}^{N} |(R\overrightarrow{p_i} + T) - \overrightarrow{q_j}|^2$$

最小时的 R 和 T，此时的 R、T 即 MVP 矩阵。其中，E为变换后两个点集中对应点的距离和，N为点集中点的个数。

步骤 9.6：得到 MVP 矩阵(Model View Projection)，进行三维图形绘制。

步骤 10：截图生成图标。

在 AR 模式的主菜单，添加一个按钮，绑定摄像机的截图功能，在菜单右侧添加滚动菜单栏，当截图函数成功触发，截图通过设置好的动态材质转换函数，显示到右侧滚动菜单栏。演示过程中，用户点击截图按键，系统会在界面左边生成图标，方便用户对学习过程中的难点、疑问点记录与详细观测，这样可以加固学习效果。

步骤 11：模型旋转停止展示。

AR 模式下，用户看到的是真实场景与虚拟模型的叠加。将设置的模型实例化为一个 Actor，添加 Rotation 函数，实现模型顺时针旋转。此设计，设置模型旋转，用户对工具模型有一个 360 度观测、学习，可以更好的达到视觉效果，这种演示学习模式更具有真实性、沉浸感。

步骤 12：AR 功能扩展模块。

本发明添加 AR 教育展示扩展功能，通过添加二级 UI，控制 Map 切换，实现不同物体的演示。其中包括地球、土星、水星、含大气层星球以及星系运行展示功能，星球做自转运动，通过 AR 模式，将运动的星球展现在用户眼前，并添加知识简介功能，完善了本系统扩展的教育展示功能。

步骤 13：动态手势与模型交互。

AR 模式添加 OpenCV 视频信息处理，初始化视频流后，先进行运动检测，如果检测到动态手运动，则进行图像处理，将手势进行图形处理去噪、转成 HSV 模式、形态学处理、画轮廓线，标定信息、轮廓信息比较，最后进行模型受力模拟，实现动态手势与虚拟模型的交互， 具体实现原理流程如图 5 所示。特别的，此动态手势交互实现了模拟三维手势的识别控制，视频流获取的动态手为二维信息，这里通过矩阵运算，将与计算得到的摄像机相对于探测到的标识的转换矩阵做比较，得到一个三维运动手势运动信息，从而实现对模型在三维空间里不同方向上的受力模拟；具体包括如下步骤：

步骤 13.1：运动检测

本方法是基于颜色直方图与背景差分的运动检测，程序在启动摄像头过程中，需要一定时间，这个时间差不多可以采集 20 帧的图像，对这 20 帧进行循环背景更新如下式，并对每

帧运动检测后得到除运动手势区域外的像素也做背景更新。

$$u_{t+1}=\begin{cases} u, & I_f(x,y)=1 \\ au_t+(1-a)I_t, & \text{其他} \end{cases}$$

其中，u_t为背景图像相应的像素点，u_{t+1}为更新后的背景图像像素点；I_t为当前帧图像的像素点，I_f是当前帧图像像素点的掩码值，即是否做背景更新；$a\in[0,1]$为背景图像模型更新速度，一般取 0.8 到 1，本方法取 0.8。

步骤 13.2：图像预处理

通过步骤 13.1 的简单运动检测步骤，如果检测到有运动信息，则开始对含有运动手势的视频帧图像进行预处理：通过 OpenCV 的 medianBlur 函数对图像进行中值滤波，去除椒盐噪声；

步骤 13.3：转换到 HSV 空间

通过 cvtColor 函数对图像进行颜色空间转换，得到其 HSV 空间的数据，并对在 HSV 空间中亮度v值重新设置为比较小的亮度值（减小静止类肤色的干扰）；对 HSV 空间中亮度v值重新设定如下式所示：

$$\begin{cases} v_{temp}=\begin{cases} r/g, & g\neq 0 \\ 4, & g=0 \end{cases} \\ v=\begin{cases} 10r/g, & v_{temp}\leq 2 \\ 255, & v_{temp}>2 \end{cases} \end{cases}$$

其中，r、g为感兴趣的肤色区域的红色与绿色像素，且$r>g$；

步骤 13.4：分割手区域，并进行形态学处理

将得到的运动二值图和通过反投影得到的二值图相与，在进行一些图像形态学闭操作得到比较完整的运动肤色手势二值图；去除图像中的杂点；

步骤 13.5：获取手势轮廓

通过初步的形态学操作，去除噪声，并使手的边界更加清晰后，通过 OpenCV 的 findContours 函数得到手势轮廓，然后进行去除伪轮廓操作；

步骤 13.6：画出轮廓，标定信息

步骤 13.7：轮廓信息比较，设置方向矢量

由于手是不断运动的，所以我们得到的轮廓也是不断在改变。将每一帧得到的轮廓进行比较，设定比较条件。通过比较给方向标志变量赋值。状态比较和分析如表 1：

表 1：状态分析

动态手势	轮廓变化	方向标记	模型动态
手掌由远到近	形状由小到大，轮廓位置中心点不变	x 轴方向矢量	向前移动
手掌由下到上	形状变化不大，轮廓位置中心点改变	z 轴方向矢量	向上移动
手掌做不同方向扭转	形状不规则改变，轮廓位置中心点不变	三维空间扭转方向矢量	根据扭转方向矢量做出相应移动

步骤 13.8：通过方向矢量，作用到虚拟模型，产生受力模拟

动态手势通过轮廓判断后，根据不同的判断结果对虚拟模型进行受力模拟操作。根据轮廓判断过程中方向标记的值，模型在三维空间的坐标值将会进行 x、y、z 三个坐标轴上的相乘计算，通过坐标值的改变，实现模型位置的改变而达到受力的模拟。

本实施例中选取一组手掌由远到近的运动、由下到上运动以及手掌向各个方向扭转运动

对模型产生的不同受力效果模拟展示，手势运动模型分别向前移动、向上移动以及根据手的不同扭转方向有一个向各方向受力的运行效果。此功能展示了动态手势与虚拟模型的交互，此交互可以帮助用户多角度观察模型，并实现了教学与用户之间的互动，增加趣味性。

当然，上述说明并非是对本发明的限制，本发明也并不仅限于上述举例，本技术领域的技术人员在本发明的实质范围内所做出的变化、改型、添加或替换，也应属于本发明的保护范围。

权 利 要 求 书

1、一种基于虚拟现实与增强现实的采矿操作多交互实现方法，其特征在于：采用井下采矿操作多交互仿真系统，该系统包含虚拟现实模式和增强现实模式两种模式；虚拟现实模式包含特定场景的建模、漫游、模型及其材质的更换、视频嵌入虚拟场景、模型移动、应用场景意向交互、二维码生成以及语音交互；增强现实模式包含模型选择、模型讲解、动态模型演示、手势控制模型交互、截图生成图标、360 度旋转以及停止、功能模式切换以及功能扩展；系统设计了两种隐藏菜单，即虚拟现实模式下的更换工具、材质的选择菜单以及增强现实模式下的模型选择类菜单；第一种用户进入特定区域菜单才会显示，离开即会隐藏；第二种点击便可在某处显示二级菜单，再次点击菜单隐藏；

所述的采矿操作多交互实现方法，具体包括如下步骤：

步骤 1：矿山开采操作的整个环境场景的搭建

根据井下采矿操作的真实环境，利用建模工具 3DMax 进行 1：1 等比建模，实现整个井下采矿操作的环境模拟；利用 UE4 引擎对模型进行包括创建、编辑贴图以及材质在内的编辑，添加物理碰撞，对整体环境进行灯光、效果光照以及特效添加，并进行烘焙、渲染；

步骤 2：虚拟现实应用场景的漫游

在 UE4 引擎中，设置键盘上、下、左、右键，绑定 Up、Down、Right、Left 方向控制函数，为鼠标绑定 Turnaround 控制函数，实现整个井下采矿操作的虚拟现实场景的漫游；

步骤 3：更换井下采矿操作的工具模型以及矿山地质的模拟材质

在虚拟井下矿山开采场景中添加隐藏菜单，当漫游至矿石开采处，会自动出现模型或者材质选择菜单，用户可以根据需求从菜单中选择模型或者材质进行更换；

步骤 4：将视频素材嵌入三维应用场景并控制播放、停止

将视频素材嵌入虚拟现实场景，在三维空间中播放，模拟矿山开采环境的监控显示设备，设置键盘 X 键，绑定 UE4 平台的 MediaPlayer 媒体类，通过 OpenSource 和 Close 函数控制视频的播放和停止；

步骤 5：选择模型并移动到任意位置

通过鼠标选中模型并将模型移动到任意需要进行模拟操作的位置，达到真实场景中的机械移动模拟；

步骤 6：实现应用场景的意向交互

当用户在虚拟现实应用场景中漫游至某一特定位置，系统检测到用户有意向进入，就自动开启环境灯，实现虚拟场景中的自然交互；

步骤 7：二维码生成

绑定键盘的 F 键，添加二维码生成函数，设置键盘按键控制生成二维码功能，用户按键盘 F 键，系统生成含设置好采样点的虚拟场景全景图的二维码；

步骤 8：实现语音交互

用户通过包括正转、反转、升臂、降臂、停止在内的关键词控制虚拟现实场景中的采煤机，模拟其运作效果；

步骤 9：AR 动态演示功能模式切换

用户点击系统右上角的 AR 模式按键切换到 AR 演示模式。

2、根据权利要求 1 所述的基于虚拟现实与增强现实的采矿操作多交互实现方法，其特征在于：在步骤 3 中，将模型实例化为具体的 Actor，添加 SetMesh 函数以及 SetMaterial 函数来更换模型和模型材质，设置 Widget Blueprint 用户界面以及 Box collision 碰撞检测，实现三维空间的隐藏菜单功能。

3、根据权利要求 1 所述的基于虚拟现实与增强现实的采矿操作多交互实现方法，其特征在于：在步骤 5 中，为要操作的模型添加鼠标事件，通过 GetHitResult 函数将模型选中，然后根据鼠标在三维空间的坐标，改变模型的 SetActorLocation 函数的坐标值，当鼠标再次点击，将此时鼠标 x、y、z 三个方向的坐标值赋给模型，此时 GetHitResult 函数将模型设置为取消选中模式。

4、根据权利要求 1 所述的基于虚拟现实与增强现实的采矿操作多交互实现方法，其特征

在于：在步骤 6 中，设置 TriggerBox 触发器，当第一人称角色触发 TriggerBox，系统检测到用户有意向进入某区域，便会自动启用此区域的某个设备。

5、根据权利要求 1 所述的基于虚拟现实与增强现实的采矿操作多交互实现方法，其特征在于：在步骤 7 中，用户按键盘 F 键，系统生成含设置好采样点的虚拟场景全景图二维码，用户用手机扫描二维码，跳转到手机端的虚拟应用场景展示页面，在手机端，用户可以启用陀螺仪，切换到 VR 分屏模式，设置好手机参数，便可用 VR 眼镜体验虚拟井下采矿操作环境场景，实现 720 度的视角展示，还能够实现手机端的多场景、多角度的漫游体验。

6、根据权利要求 1 所述的基于虚拟现实与增强现实的采矿操作多交互实现方法，其特征在于：在步骤 8 中，语音识别基于 Pocket-sphinx 库实现，通过改进中文关键字字典，经过预处理、特征提取、声学模型训练、语言模型训练以及语音解码和搜索实现识别功能，最后经过 UE4 引擎的编写功能控制函数实现三维空间里语音对模型的控制；语音识别的具体实现步骤如下：

步骤 8.1：预处理

对输入的原始语音信号进行处理，滤除掉其中的不重要的信息以及背景噪声，并对语音信号的端点检测、语音分帧和预加重进行处理；

通过一阶 FIR 高通数字滤波器来实现预加重，一阶 FIR 高通数字滤波器的传递函数为：

$$\mathbf{H}(\boldsymbol{z}) = \mathbf{1} - \boldsymbol{a}\boldsymbol{z}^{-\mathbf{1}};$$

其中，a 为预加重滤波器的系数，取值范围为 0.9~1.0，若设 n 时刻的语音采样值为 x(n)，则预加重后的信号为

$$\mathrm{y(n)} = \mathrm{x(n)} - \mathrm{a} * \mathrm{x(n-1)};$$

步骤 8.2：特征提取

通过梅尔频率倒谱系数（MFCC）的方法来进行特征提取；具体按照如下步骤进行：

步骤 8.2.1：利用人听觉的临界带效应，采用 MEL 倒谱分析技术对语音信号处理得到 MEL 倒谱系数矢量序列；

步骤 8.2.2：用 MEL 倒谱系数矢量序列表示输入语音的频谱，在语音频谱范围内设置若干个具有三角形或正弦形滤波特性的带通滤波器；

步骤 8.2.3：通过带通滤波器组，求各个带通滤波器的输出数据；

步骤 8.2.4：对各个带通滤波器的输出数据取对数，并做离散余弦变换(DCT)；

步骤 8.2.5：得到 MFCC 系数；求解公式如下：

$$\boldsymbol{C_i} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{P}} \mathbf{log}\boldsymbol{F(k)}\,\mathbf{cos}[\boldsymbol{\pi(k-0.5)i/P}] \qquad \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots \boldsymbol{P};$$

其中，$\boldsymbol{C_i}$为特征参数，$\boldsymbol{k}$为三角滤波器的个数，F(k)为各个滤波器的输出数据，P为滤波器阶数，$\boldsymbol{i}$为数据长度；

步骤 8.3：声学模型训练

根据训练语音库的特征参数训练出声学模型参数；

在识别时可以将待识别的语音的特征参数同声学模型进行匹配，得到识别结果；本文采用混合高斯模型-隐马尔科夫模型（GMM-HMM）作为声学模型，具体包括如下步骤：

步骤 8.3.1：求出混合高斯模型的联合概率密度函数的形式如下：

$$\mathrm{P(x)} = \sum_{m=1}^{M} \frac{C_m}{(2\pi)^{D/2}|\sum_m|^{1/2}} exp\left[-\frac{1}{2}(x-u_m)^T \sum_m^{-1}(x-u_m)\right]$$

$$= \sum_{\boldsymbol{m=1}}^{\boldsymbol{M}} C_m\, \boldsymbol{N}(\boldsymbol{x}; \boldsymbol{u_m}, \boldsymbol{\sum_m})$$

其中，M表示混合高斯模型中高斯的个数，C_m表示权重，$\boldsymbol{u_m}$表示均值，$\boldsymbol{\sum_m}$表示协方差

矩阵，D 为观测矢量维数；利用最大期望值算法（EM）对混合高斯模型参数变量 $\Theta = \{C_m, u_m, \sum_m\}$进行估计，利用如下公式求解：

$$c_m^{(j+1)} = \frac{1}{N}\sum_{t=1}^{N} h_m^{(j)}(t)$$

$$u_m^{(j+1)} = \frac{\sum_{t=1}^{N} h_m^{(j)}(t)x^{(t)}}{\sum_{t=1}^{N} h_m^{(j)}(t)}$$

$$\sum_m^{(j+1)} = \frac{\sum_{t=1}^{N} h_m^{(j)}\left[x^{(t)} - u_m^{(j)}\right]\left[x^{(t)} - u_m^{(j)}\right]^T}{\sum_{t=1}^{N} h_m^{(j)}(t)}$$

其中，j是当前迭代轮数，N表示训练数据集中元素的个数，$x^{(t)}$为t时刻的特征向量，$h_m(t)$表示 t时刻C_m的后验概率；GMM 参数通过 EM 算法进行估计，可以使其在训练数据上生成语音观察特征的概率最大化；

步骤 8.3.2：求解 HMM 的三个组成部分

设状态序列为$q_1, q_2, \ldots, q_N$，令转移概率矩阵$A = [a_{ij}]\ i, j \in [1, N]$，则求出的马尔科夫链状态间的跳转概率为：$a_{ij} = P(q_t = j | q_{t-1} = i)$；马尔科夫链的初始概率$\pi = [\pi_i]\ i \in [1, N]$，其中，$\pi_i = P(q_1 = i)$；令每个状态的观察概率分布$b_i(o_t) = P(o_t | q_t = i)$，采用 GMM 模型来描述状态的观察概率分布；根据步骤 8.3.1，求解公式为：

$$b_i(o_t) = \sum_{m=1}^{M} \frac{C_{i,m}}{(2\pi)^{D/2}|\sum_{i,m}|^{1/2}} exp\left[-\frac{1}{2}(o_t - u_{i,m})^T \sum_{i,m}^{-1}(o_t - u_{i,m})\right]$$

其中，N为状态个数，i、j表示状态，a_{ij}表示$t-1$时刻从i状态跳转到t时刻j状态的概率，o_t为t时刻的观测值，$C_{i,m}$为混合系数，表示不同高斯之间的权重，$u_{i,m}$表示不同高斯之间的均值，$\sum_{i,m}$表示不同高斯之间的协方差矩阵；HMM 的参数通过 Baum-Welch 算法进行估计得出，最后生成声学模型文件；

步骤 8.4：语言模型训练

采用 N-Gram 模型实现语言模型的训练；在一个语句中第i个词出现的概率，条件依赖于它前面的 N-1 个词，即将一个词的上下文定义为该词前面出现的 N-1 个词，其表达式为：

$$\mathrm{P}(w_i = \mathrm{w}|\mathrm{c}) = \mathrm{P}(w_i = \mathrm{w}|w_{i-N+1}^{i-1})$$

使用条件概率公式 S 将上述表达式替换成如下公式：

$$\mathrm{P}(\mathrm{sentence}) = P(w_1) * P(w_2|w_1) * P(w_3|w_2) \ldots * P(w_n|w_1, w_2, \ldots, w_{n-1})$$

其中，$\mathrm{P}(w_1)$是w_1在文章中出现的概率，$\mathrm{P}(w_1, w_2)$是w_1, w_2连续出现的概率，$\mathrm{P}(w_2|w_1)$是已知w_1已出现的情况下出现w_2的概率，假设识别 sentence 的概率用$\mathrm{P}(s)$表示，$\mathrm{P}(s) = P(w_1, w_2, \ldots, w_n)$ 表示单词集$w_1, w_2, \ldots, w_n$连续出现并生成 S 的概率；

通过马尔科夫假设精简成如下公式：

$$\mathrm{P}(\mathrm{sentence}) = P(w_1) * P(w_2|w_1) * P(w_3|w_2) \ldots * P(w_n|w_{n-1})$$

其中，$P(w_i|w_{i-1}) = P(w_{i-1}, w_i)/P(w_i)$，$P(w_{i-1}, w_i)$和$P(w_i)$都可以从语料统计出来，最终就能得到P(sentence)，语言模型存储$P(w_{i-1}, w_i)$的概率统计值，通过求出 P(sentence)的最大值来实现整个识别过程；

步骤 8.5：语音解码和搜索算法

针对输入的语音信号，根据已经训练好的声学模型、语言模型及利用 g2p 工具创建好的字典映射文件建立一个识别网络，根据搜索算法在该网络中寻找最佳的一条路径，这个路径就是能够以最大概率输出该语音信号的词串，这样就确定这个语音样本所包含的文字，本文采用 Viterbi 算法实现语音解码，具体过程如下：

步骤 8.5.1：输入 HMM 模型的参数和观测序列$O = \{o_1, o_2, \ldots, o_T\}$，则 t=1 时所有的状态概率为：

$$\delta_1(i) = \pi_i b_i(o_1)$$

$$\psi_1(i) = 0$$

步骤 8.5.2：逐渐递推到 t=2,3,…,T，则为：

$$\delta_t(i) = \max_{1 \le j \le N} [\delta_{t-1} a_{ij}] b_i(o_t)$$

$$\psi_t(i) = \operatorname{argmax}_{1 \le j \le N} [\delta_{t-1}(j) a_{ji}]$$

步骤 8.5.3：终止遍历：

$$P^* = \max_{1 \le i \le N} \delta_T(i)$$

$$i_T^* = \operatorname{argmax}_{1 \le i \le N} [\delta_T(i)]$$

步骤 8.5.4：回溯最优路径，t=T-1,T-2,…,1；

$$i_t^* = \psi_{t+1}(i_{t+1}^*)$$

步骤 8.5.5：输出最优隐状态路径$I^* = (i_1^*, i_2^*, \ldots, i_T^*)$；

其中，$\delta_t(i)$是递推到 t 时刻，最优路径经过的所有结点的联合概率，$\psi_t(i)$是 t 时刻的隐状态，T为时间，P^*为最优路径的概率，i_T^*为最优路径的终结点。

7、根据权利要求 6 所述的基于虚拟现实与增强现实的采矿操作多交互实现方法，其特征在于：a 取 0.97。

8、根据权利要求 1 所述的基于虚拟现实与增强现实的采矿操作多交互实现方法，其特征在于：在步骤 9 中，具体包括如下步骤：

步骤 9.1：模型选择

对采煤机模型、掘进机模型、风煤钻模型以及综采支架模型进行选择，每一类模型都是对真实采煤工具的 1:1 建模模拟；

步骤 9.2：模型讲解

用户通过选择模型后，再通过此菜单选择需要学习的工具模型选项，系统会播放对应的语音讲解，再次点击按键语音停止；

步骤 9.3：模型演示

将在 3DMax 建模过程中制作的工具模拟运行动画导入 UE4 引擎中，设置相应的选择菜单，点击便可在 AR 模式下演示相应采煤工具的运行状态；

步骤 9.4：截图生成图标

在 AR 模式的主菜单，添加一个按钮，绑定摄像机的截图功能，在菜单右侧添加滚动菜单栏，当截图函数成功触发，截图通过设置好的动态材质转换函数，显示到右侧滚动菜单栏，演示过程中，用户点击截图按键，系统会在界面一侧生成图标；

步骤 9.5：旋转

将设置的模型实例化为一个 Actor，添加 Rotation 函数，实现模型顺时针旋转；

步骤 9.6：功能扩展

添加二级 UI，控制 Map 切换，实现包括地球、土星、水星、含大气层星球以及星系在内的运行演示功能；添加 WidgetBlueprint 编码实现了知识简介面板的显示或隐藏；设计返回键回到 AR 编辑主模块；

步骤 9.7：动态手势控制模型，真实环境与虚拟模型叠加，手势与模型进行交互控制，具体包括如下步骤：

步骤 9.7.1：初始化视频捕捉，读取标志文件和摄像相机参数；

步骤 9.7.2：抓取视频帧图像；

步骤 9.7.3：执行探测标记以及识别视频帧中的标记模板，并利用 OpenCV 库函数对获取的视频帧图像进行运动检测，判断是否检测到运动轨迹；

若：判断结果是检测到手势运动轨迹，则执行步骤 9.7.4；

或判断结果是没有检测到运动轨迹，则继续执行探测标记以及识别视频帧中的标记模板，然后执行步骤 9.7.12；

基于颜色直方图与背景差分进行运动检测，对采集的帧以及对每帧运动检测后得到除运动手势区域外的像素做背景更新，公式如下；

$$u_{t+1}=\begin{cases} u, & I_f(x,y)=1 \\ au_t+(1-a)I_t, & \text{其他} \end{cases}$$

其中，u_t为背景图像相应的像素点，u_{t+1}为更新后的背景图像像素点；I_t为当前帧图像的像素点，I_f是当前帧图像像素点的掩码值，即是否做背景更新；$a \in [0,1]$为背景图像模型更新速度，本文取 0.8；

步骤 9.7.4：对图像进行包括去噪在内的预处理；

通过运动检测步骤，如果检测到有运动信息，则开始对含有运动手势的视频帧图像进行预处理：通过 OpenCV 的 medianBlur 函数对图像进行中值滤波，去除椒盐噪声；

步骤 9.7.5：转换到 HSV 空间；

通过 cvtColor 函数对图像进行颜色空间转换，得到其 HSV 空间的数据，并对 HSV 空间中的亮度v值重新设定如下式所示：

$$\begin{cases} v_{temp}=\begin{cases} r/g, & g\neq 0 \\ 4, & g=0 \end{cases} \\ v=\begin{cases} 10r/g, & v_{temp}\leq 2 \\ 255, & v_{temp}>2 \end{cases} \end{cases}$$

其中，r、g为肤色区域的红色与绿色像素，且$r>g$；

步骤 9.7.6：分割手区域；

步骤 9.7.7：进行形态学处理，去除杂点；

将得到的运动二值图和通过反投影得到的二值图相与，并进行图像形态学闭操作得到比较完整的运动肤色手势二值图；并去除图像中的杂点；

步骤 9.7.8：获取手轮廓；

通过初步的形态学操作，去除噪声，并使手的边界更加清晰后，通过 OpenCV 的 findContours 函数得到手势轮廓，然后进行去除伪轮廓操作；

步骤 9.7.9：画出手轮廓，标定信息；

步骤 9.7.10：轮廓信息比较，设置方向向量；

将每一帧得到的轮廓进行比较，设定比较条件，通过比较给方向标志变量赋值；

步骤 9.7.11：对模型根据矢量坐标进行受力模拟，实现动态手势与虚拟模型的交互；

动态手势通过轮廓判断后，根据不同的判断结果对虚拟模型进行受力模拟操作，根据轮

廓判断过程中方向标记的值，模型在三维空间的坐标值将会进行 x、y、z 三个坐标轴上的相乘计算，通过坐标值的改变，实现模型位置的改变而达到受力的模拟；

步骤 9.7.12：计算摄像头相对于探测到的标记的转换矩阵；

步骤 9.7.13：在探测到的标记上叠加虚拟物体，并返回执行步骤 9.7.2，实现真实环境与虚拟模型的叠加显示；

步骤 9.7.14：当点击 VR 模式时，系统切换显示模式，摄像头关闭，以上步骤停止执行。

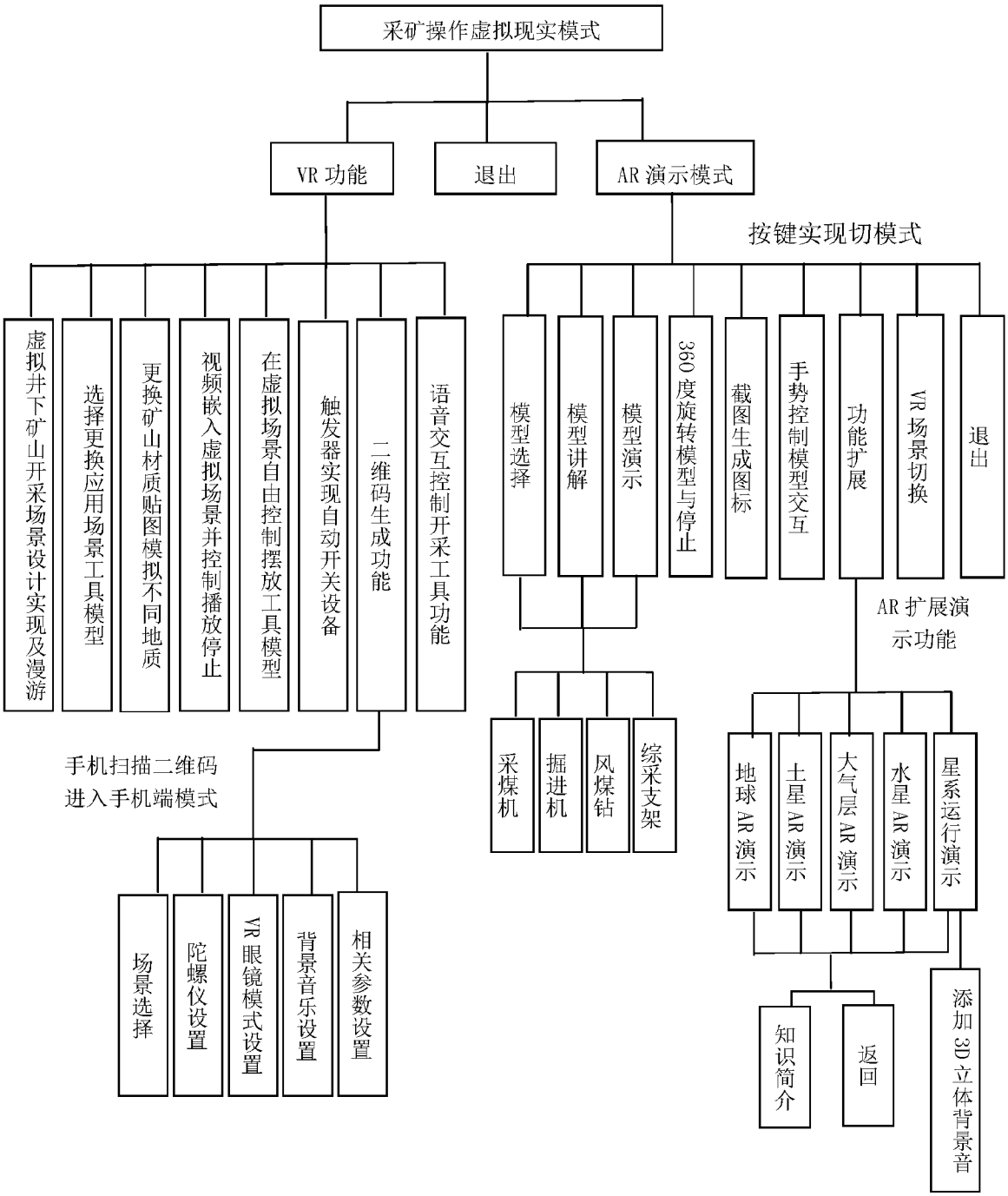
采矿操作虚拟现实模式
VR 功能
退出
AR 演示模式
按键实现切模式
虚拟井下矿山开采场景设计实现及漫游
选择更换应用场景工具模型
更换矿山材质贴图模拟不同地质
视频嵌入虚拟场景并控制播放停止
在虚拟场景自由控制摆放工具模型
触发器实现自动开关设备
二维码生成功能
语音交互控制开采工具功能
模型选择
模型讲解
模型演示
360 度旋转模型与停止
截图生成图标
手势控制模型交互
功能扩展
VR 场景切换
退出
AR 扩展演示功能
采煤机
掘进机
风煤钻
综采支架
地球 AR 演示
土星 AR 演示
大气层 AR 演示
水星 AR 演示
星系运行演示
知识简介
返回
添加 3D 立体背景音
手机扫描二维码
进入手机端模式
场景选择
陀螺仪设置
VR 眼镜模式设置
背景音乐设置
相关参数设置

图 1

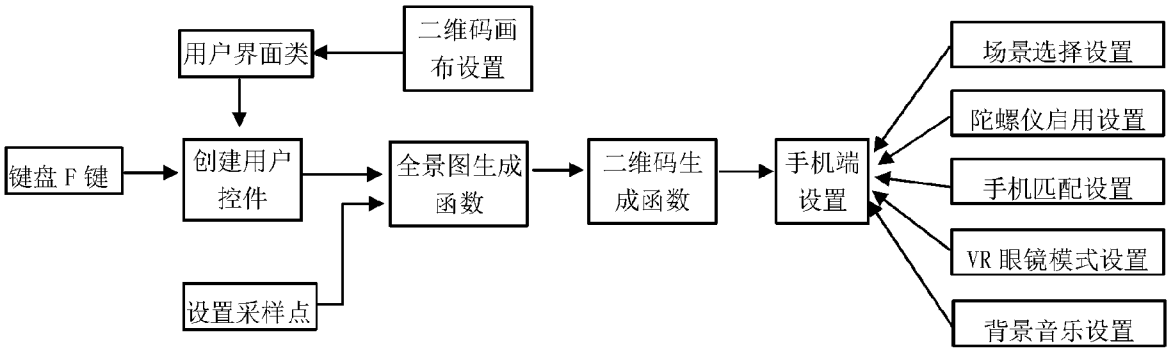
用户界面类
二维码画布设置
键盘 F 键
创建用户控件
全景图生成函数
二维码生成函数
手机端设置
场景选择设置
陀螺仪启用设置
手机匹配设置
VR 眼镜模式设置
背景音乐设置
设置采样点

图 2

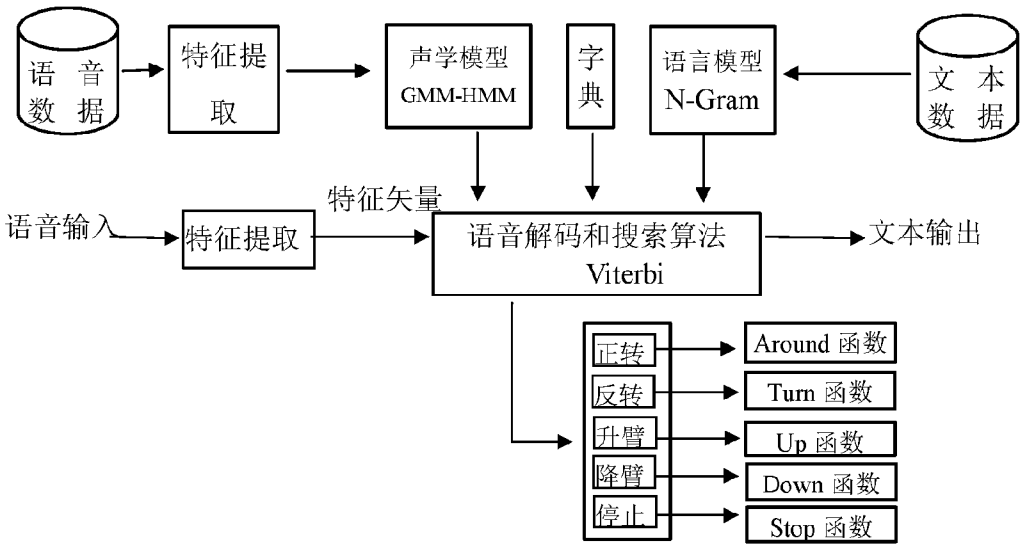
语音数据
特征提取
声学模型 GMM-HMM
字典
语言模型 N-Gram
文本数据
语音输入
特征提取
特征矢量
语音解码和搜索算法 Viterbi
文本输出
正转
反转
升臂
降臂
停止
Around 函数
Turn 函数
Up 函数
Down 函数
Stop 函数

图 3

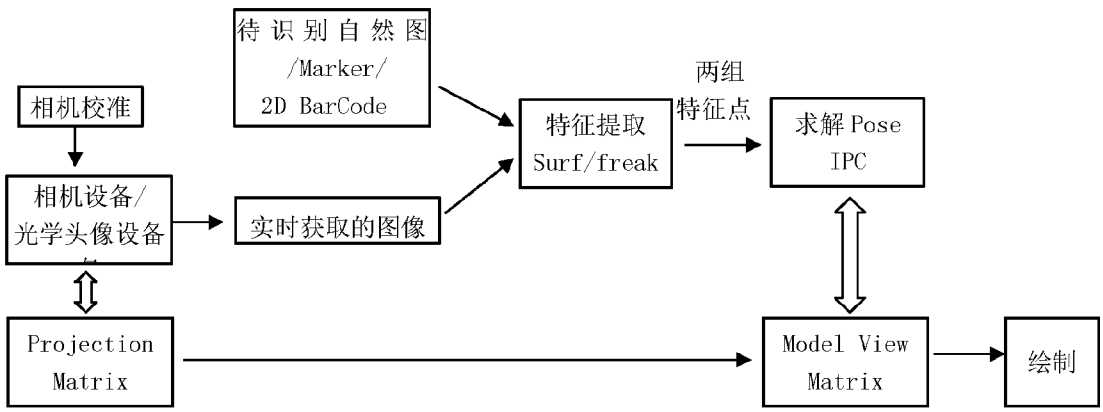
待识别自然图/Marker/2D BarCode
相机校准
相机设备/光学头像设备
实时获取的图像
特征提取 Surf/freak
两组特征点
求解 Pose IPC
Projection Matrix
Model View Matrix
绘制

图 4

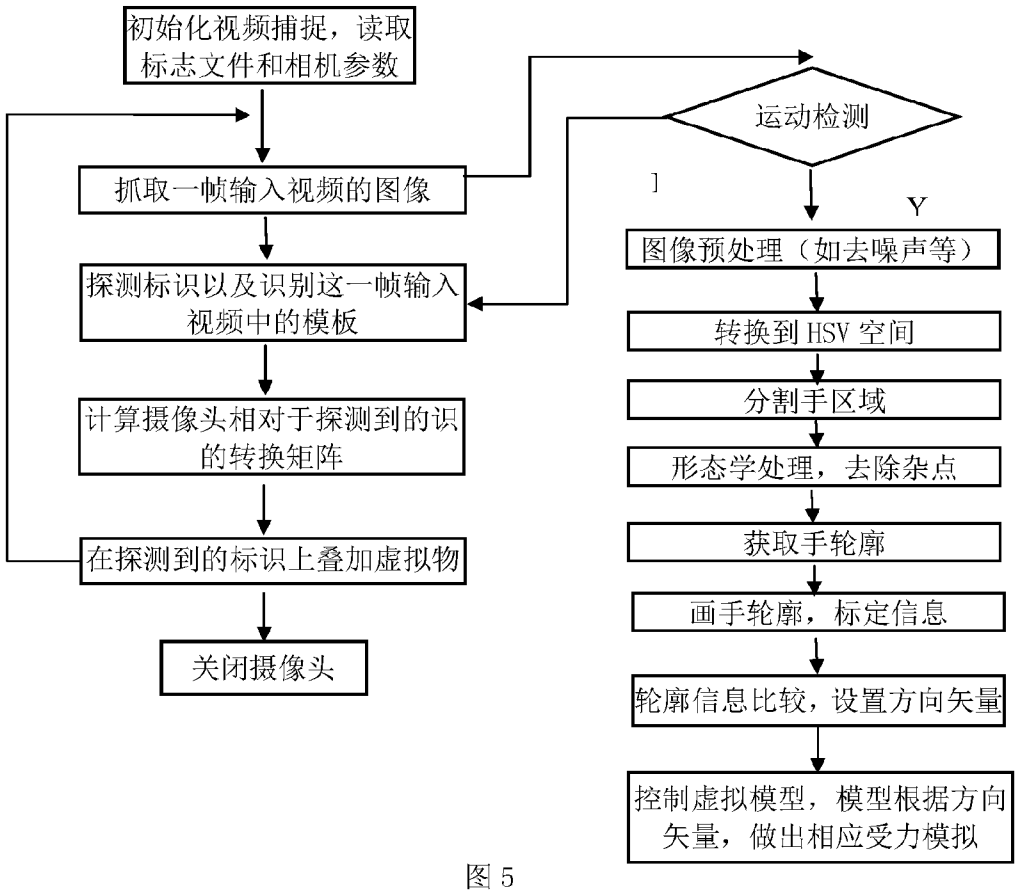
初始化视频捕捉，读取标志文件和相机参数
抓取一帧输入视频的图像
探测标识以及识别这一帧输入视频中的模板
计算摄像头相对于探测到的识的转换矩阵
在探测到的标识上叠加虚拟物
关闭摄像头
运动检测
]
Y
图像预处理（如去噪声等）
转换到 HSV 空间
分割手区域
形态学处理，去除杂点
获取手轮廓
画手轮廓，标定信息
轮廓信息比较，设置方向矢量
控制虚拟模型，模型根据方向矢量，做出相应受力模拟

图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/118923

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

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)

WPI, EPODOC, CNKI, CNPAT, IEEE, GOOGLE: 虚拟现实, 增强现实, 环境, 采矿, 交互, 切换, 模式, VR, AR, environment, mining, interact+, switch+, mode

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 107515674 A (UNIVERSITY OF SCIENCE AND TECHNOLOGY OF SHANDONG) 26 December 2017 (26.12.2017), the abstract, description, paragraphs [0005]-[0131], and figures 1-5	1-8
Y	CN 105955456 A (SHENZHEN SUPER PERFECT OPTOELECTRONICS CO., LTD.) 21 September 2016 (21.09.2016), the abstract, description, paragraphs [0073]-[0091], and figures 1-5	1-8
Y	CN 106019364 A (DALIAN UNIVERSITY OF TECHNOLOGY et al.) 12 October 2016 (12.10.2016), the abstract, description, paragraphs [0009] and [0019], and figure 1	1-8
A	US 2016090839 A1 (STOLARCZYK, LARRY G.) 31 March 2016 (31.03.2016), entire document	1-8

☐ 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
07 March 2018	27 March 2018

Name and mailing address of the ISA
State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao
Haidian District, Beijing 100088, China
Facsimile No. (86-10) 62019451

Authorized officer
WANG, Yanchen
Telephone No. (86-10) 53961435

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/118923

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 107515674 A	26 December 2017	None	
CN 105955456 A	21 September 2016	US 2017301137 A1	19 October 2017
CN 106019364 A	12 October 2016	None	
US 2016090839 A1	31 March 2016	None	

国际检索报告

国际申请号
PCT/CN2017/118923

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

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））

WPI, EPODOC, CNKI, CNPAT, IEEE, GOOGLE: 虚拟现实, 增强现实, 环境, 采矿, 交互,切换, 模式, VR, AR, environment, mining, interact+, switch+, mode

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 107515674 A (山东科技大学) 2017年 12月 26日 (2017-12-26) 摘要、说明书第0005-0131段、附图1-5	1-8
Y	CN 105955456 A (深圳超多维光电子有限公司) 2016年 9月 21日 (2016-09-21) 摘要、说明书第0073-0091段、附图1-5	1-8
Y	CN 106019364 A (大连理工大学 等) 2016年 10月 12日 (2016-10-12) 摘要、说明书第0009, 0019段、附图1	1-8
A	US 2016090839 A1 (STOLARCZYK, LARRY G.) 2016年 3月 31日 (2016-03-31) 全文	1-8

☐ 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
2018年 3月 7日	2018年 3月 27日

ISA/CN的名称和邮寄地址	受权官员
中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	王艳臣
传真号 (86-10)62019451	电话号码 (86-10)01053961435

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2017/118923

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 107515674 A	2017年 12月 26日	无	
CN 105955456 A	2016年 9月 21日	US 2017301137 A1	2017年 10月 19日
CN 106019364 A	2016年 10月 12日	无	
US 2016090839 A1	2016年 3月 31日	无	