

doi: 10.7690/bgzdh.2023.10.013

基于增强现实的视力自查互动系统

张兆灿, 李淑琴

(北京信息科技大学计算机学院, 北京 100101)

摘要: 为利用增强现实(augmented reality, AR)技术实现视力自查互动系统,使用前端框架、canvas 2d 渲染引擎和 TensorFlow.JS 等技术,并设计系统架构和互动测试视力流程,设计实现模型训练页面、互动测试 UI 部分及对应的动画元素。通过摄像头捕获实时的人物动作,实现系统基于 AR 的简单互动,系统可以识别用户的表情与手势组合符合哪种预先设定组合,并做出字母方向识别正确与否的判断;通过屏显设备在真实视力测试的场景内,实现 AR 互动式测试视力的效果,经评估该系统在拥有良好用户体验的情况下实现准确判断。结果表明:通过探究 AR 技术实现视力自查互动系统,可使用户随时随地自查视力来爱护眼睛,保护视力,并取得了良好的效果。

关键词: AR; 视力自查互动系统; AR 互动; 实时动作捕捉

中图分类号: TP391.4 **文献标志码:** A

Vision Self-examination Interactive System Based on Augmented Reality

Zhang Taocan, Li Shuqin

(School of Computer, Beijing Information Science and Technology University, Beijing 100101, China)

Abstract: In order to use augmented reality (AR) technology to realize vision self-examination interactive system, the front-end framework, canvas 2D rendering engine and TensorFlow. JS technology are used, and the system architecture and interactive vision test process are designed. Design and implement the model training page, interactive test UI part and corresponding animation elements. Capturing real-time character actions through a camera to realize simple interaction of the system based on AR, wherein the system can identify which preset combination the expression and gesture combination of the user conforms to, and judge whether the letter direction recognition is correct or not; Through the screen display device in the real vision test scene, the effect of AR interactive vision test is realized, and the system is evaluated to achieve accurate judgment under the condition of good user experience. The results show that through the exploration of AR technology to achieve vision self-examination interactive system, users can check their eyesight at any time and anywhere to care for their eyes and protect their eyesight, and achieve good results.

Keywords: AR; interactive vision self-examination system; AR interaction; real-time action capture

0 引言

增强现实(AR)是以3维注册、虚实结合、实时人机交互为特征^[1],将虚拟物体或计算机生成的其他信息叠加在现实世界中,从而实现现实“增强”的技术^[2]。从1990年首次提出“augmented reality”至今,AR已经从最初的话题和畅想,走到如今技术相对成熟、应用不断涌现的阶段^[3],国内外开发者从未停止过探索的脚步^[4-5]。

随着电子产品的兴起,几乎家家户户都有智能手机和电脑设备。不管是大人还是孩子,办公还是学习都离不开电子产品,确实方便了人们的生活,但随着人们的目光停留在屏幕上的时间变得越来越长,给眼睛带来了健康隐患。

笔者利用AR技术实现一个视力自查互动系统,在真实视力测试的场景内使用AR,呼吁人们

关注眼部健康,对于AR在生活健康的未来研究具有现实意义。

1 视力自查互动系统总体设计

1.1 视力自查互动系统整体设计

系统着重关注AR系统界面的设计实现和AR系统的关键指标,即如何将增强功能与现实世界互动需求整合在一起这2方面内容。根据情景设计、交互设计和视觉设计的思想,实现简化AR应用的交互体验,并达到视力测试简单快捷的效果。

1.1.1 视力自查互动系统情景设计

系统实现策略如图1所示,使用摄像头实时捕捉人物的动作,让人物能在AR体验中输入运动和姿势,成为AR体验的焦点。借助AR的特点,利用摄像头捕捉到的画面来进行手势或人体姿态识别

并结合视力检查。

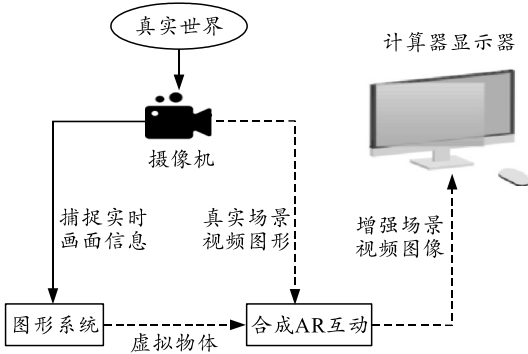


图 1 Monitor-based 增强现实系统实现方案

1.1.2 视力自查互动系统交互设计

AR 体验是叠加在真实环境之上，具有空间化和高度情境化特点。为实现良好的 AR 体验，系统设计考虑用户需求和使交互流程尽可能便捷。

1.1.3 视力自查互动系统视觉设计

系统遵照国际视力表标准，采用“E”作为视标。设计考虑显示界面符合人物测试视力的日常操作习惯，充分利用显示区域醒目地突出重要信息，也酌情考虑不同因素对操作者的影响，进行合理的规划设计。

1.2 系统功能结构设计

基于当前的系统功能需求，系统的架构设计重点从顶层至底层进行了拆分。系统架构及技术实现方案如图 2 所示，自上而下分为 UI 层、互动层和算法层。

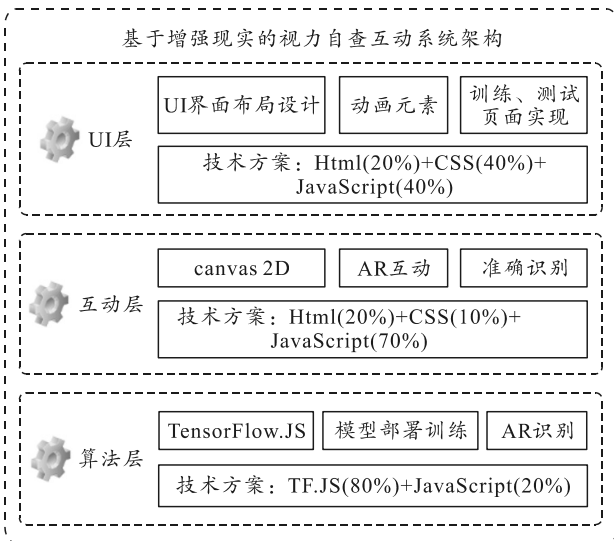


图 2 视力自查互动系统架构

1.3 系统运行总体流程

AR 自查视力的互动检测流程如图 3 所示。

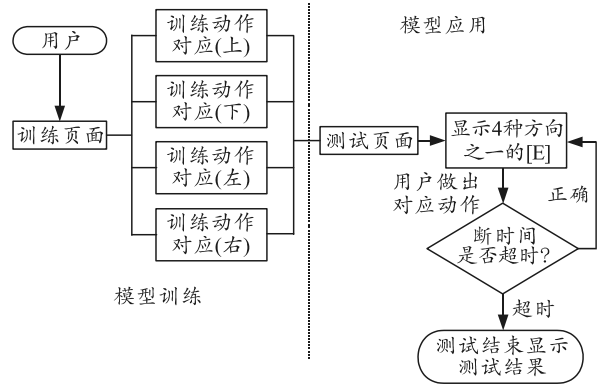


图 3 互动系统流程

- 1) 用户先在训练页面根据上下左右姿势进行训练；
- 2) 训练完成后在测试页面根据显示的 E 的方向，需做出对应的姿势；
- 3) 测试直到用户做出错误姿势或超时为止；
- 4) 根据用户的成绩会得到相应的结果。

2 视力自查互动系统 UI 层界面设计与实现

视力自查互动系统能适应智能化需求，实现视力自测且具有测试操作便捷简单的特点。视力自查互动系统的界面构成，按照模块化分为视力测试主界面和训练互动界面。

2.1 视力测试主界面设计与实现

视力测试主界面的实现模拟了 Mac Dock 菜单栏的设计，使用基于 HTML 5 的 3D 无限循环图片流来对视标进行类似旋转画廊的效果展示。画廊中的视标围成一个圆环，拖拽鼠标或点击左右转换箭头可以让不同视标在环形画廊中滚动切换，形成一种 3D 的视差效果，并对每个图片流设置倒影，增添 3D 立体的视觉效果和趣味。使用渐变色对背景颜色进行渲染，效果如图 4 所示。

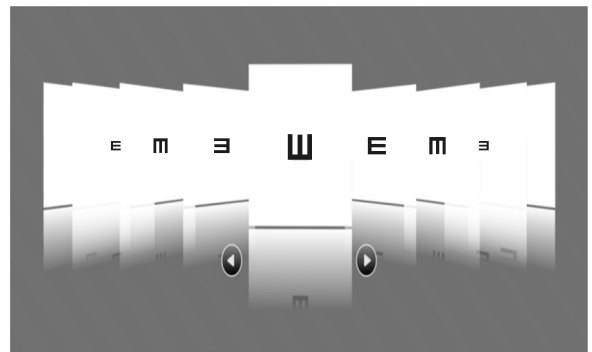


图 4 视力测试主页面

2.2 训练互动页面设计与实现

根据系统的 AR 识别及互动需求，训练互动页

面设计了控制面板。主要有上、下、左、右 4 种视频截图的姿态框来实现视频流中视频帧的图像展示、主要的操作按钮 TARIN MODEL 和 PLAY、训练参数的设定选项，且以 3D 立体动态来呈现。最终布局效果实现如图 5 所示。

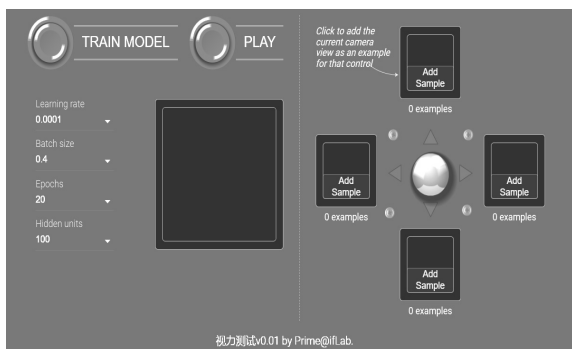


图 5 训练互动页面效果

3 视力自查互动系统互动层设计与实现

3.1 AR 互动获取视频流

系统的 AR 识别需要借助摄像头进行视频流的捕捉，调用 JavaScript 的 MediaStream API 捕获由摄像机生成的视频流，具体实现：

- 1) 调用设备摄像头；
- 2) <video>标签获取视频流，显示摄像头图像；
- 3) 将摄像头捕获的图像实时渲染放置在页面<canvas>标签；
- 4) 读取视频帧绘制 4 个方向上的预览缩略图。

3.2 读取视频帧绘制预览缩略图

捕获视频帧做法是让代码能够不断地收到从 camera 传过来的每一帧的效果，将获取到的媒体流置入 video 标签，具体的实现逻辑步骤：

1) 场景中创建 camera 对象和 canvas 对象，camera 对象用来实时获取显示场景中的图像，把每一帧的图像实时传给后台模型进行处理，处理的结果会在 canvas 对象显示。

2) 利用 CSS 调整 camera 的样式并进行相关配置，比如设定分辨率及大小，设定使用前置摄像头，关闭闪光灯等。

3) 添加按钮并设置按钮的相应功能，该按钮可实现从画布动态创建图像并将其添加到页面。

4) 当 UI 按钮被按下时，从网络摄像头读取一帧并将其与按钮给出的类标签相关联，实现了互动层 AR 互动的的基本效果。

在视频加载之前添加了 loading 加载动画，效果如图 6 所示动态旋转画面。

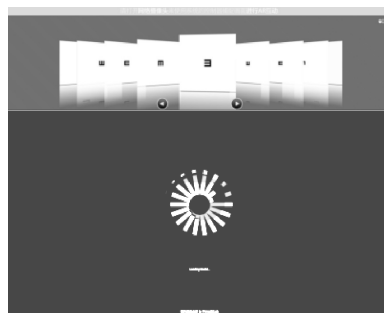


图 6 Loading 加载动画元素

4 视力自查互动系统算法模型设计与实现

4.1 算法层的设计

系统算法层需要实现的需求是：从网络摄像头数据(可以是姿势、物体、面部表情等)预测用户定义的上、下、左、右 4 个类别。

训练系统的模型，首先需要实现网络架构，但由于没有足够的训练数据，所以利用迁移学习技术来实现加快训练速度与减少训练数据的效果。通过重用预训练模型的较底层，替换掉原始模型的输出层，使用预训练模型 MobileNet^[6]的内部激活来获取特征提取能力，以此来获得良好的训练效果。

4.2 模型部署与训练

4.2.1 获取张量数据

在训练模型前，需要实现从网络摄像头获取张量。在代码中实现类去从<video>标签中读取图像并返回 TensorFlow.JS 的张量，具体实现步骤如下：

1) 从网络摄像头控件的<video>元素读取一帧，并返回形状为[高度, 宽度, 3]的张量。其中的维度 3 是对应 RGB 通道。

2) 预处理图像。摄像头输送的数据长宽比是矩形的，而 mobilenet_v1_0.25_224 模型需要 224×224 大小的方形的输入图像，需要进行图像预处理，对摄像头的 video 元素中截取视频帧的图像的高度与宽度规定为 224×224 的方形块。

3) 将预处理后的图像与按钮给出的类标签相关联，上、下、左、右分别是数字标签 0、1、2、3。

4.2.2 加载预训练模型

系统依赖 TensorFlow.JS 实现算法在浏览器中的部署，使用的预训练模型是由 Google 开发的 mobilenet_v1_0.25_224 网络结构，已在 ImageNet 数据集上进行过预训练，具有强大的特征提取能力。使用 TF 库的 loadLayersModel()方法加载预训练模型到网页中。

4.2.3 构建模型拓扑结构

对预训练模型进行修改,构建模型的拓扑结构,使模型适应新的分类任务。具体的实现步骤:

1) 创建一个 2 层的全连接模型。模型的第一层是一个 flatten 层,将输入扁平化为向量,以便可以在密集层中使用它。

2) 将 flatten 层的 inputShape 参数设置与截断 MobileNet 的激活一致。

3) 添加密集层,它的参数设置如下:

units: ui.getDenseUnits(), activation: 'relu', kernelInitializer: 'varianceScaling', useBias: true。
其中,units 是用户从 UI 中选择的单元数量,添加 relu 激活函数,内核初始化器是 varianceScaling,并添加偏差项。

4) 最后一层为另一个密集层,使用 softmax 激活函数,单元数与系统需要预测的类数相对应,即 units: NUM_CLASSES。

5) 创建驱动模型训练的优化器,使用 categoricalCrossentropy 分类交叉熵作为分类的损失函数,它测量预测的类概率分布(输入属于每个类的概率)与标签(真实类中的概率 100%)之间的误差。

5 视力自查互动系统效果展示

基于 AR 的视力自查互动检测系统主要如图 7 所示,由收集数据、训练神经网络和 AR 互动视力测试 3 部分构成;因此,运行效果也通过这 3 个阶段依次进行展示。

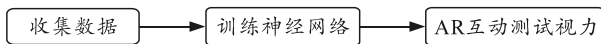


图 7 系统运行总体流程

5.1 收集数据展示

首先进行数据收集阶段。用户需要在系统弹出的允许系统访问设备摄像头弹窗中点击允许按钮来捕获视频流,如图 8 所示。

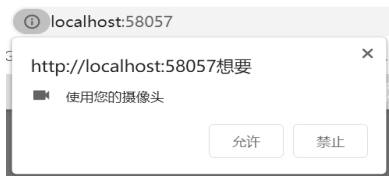


图 8 调用设备 API 请求访问摄像头

然后,用户点击训练页面(图 9 左侧)中的上、下、左、右 4 个方向控件上名为“Add Sample”的 UI 按钮,在 4 个方向上就有对应的图像显示,其效果类似对视频中的画面进行了屏幕快照的操作。同时视频显示框(图 9 中部)中同步进行着前置摄像头

的视频动画的画面捕捉显示。整体效果界面如图 9 所示,每个方向上都收集到标签和图像。

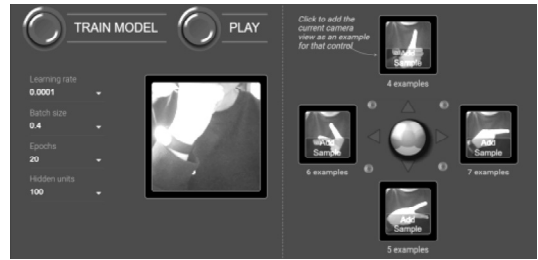


图 9 捕捉视频帧效果

5.2 训练神经网络展示

接下来进行模型训练,点击训练页面(图 9 左上角)的 TRAIN MODEL 按钮开始进行模型训练,也可根据训练效果对参数进行调整再训练,训练完成后就可以用这个神经网络来预测用户定义的 4 个类,如图 10 所示得到训练的结果 LOSS 值为 0.010 30,达到了较好的训练效果。



图 10 训练模型完成后效果

5.3 视力测试展示

用户点击训练页面上 PLAY 按钮即可通过训练好的模型进行视力测试。系统显示 4 种方向之一的视标,用户做出对应判断姿态,系统使用训练过的模型,通过网络摄像头的数据进行上、下、左、右 4 个方向的预测,并将它们输入到视力测试的过程中。在训练页面上点亮显示对应动作的方向来完成视力互动测试。如图 11 所示,显示视标方向朝下,但是捕捉画面用户手势指向左对应姿态定义为“上”,表示错误。

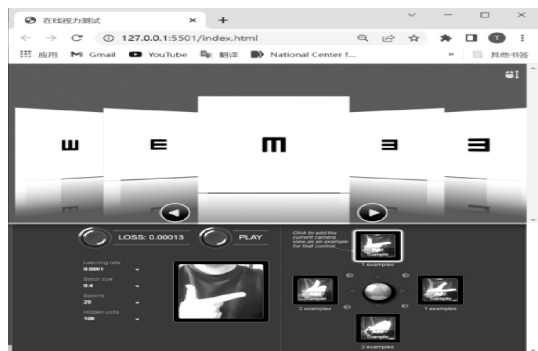


图 11 系统运行示意图 1

如图 12 所示, 显示视标方向朝下, 捕捉画面用户手势对应姿态定义为“下”, 表示正确。

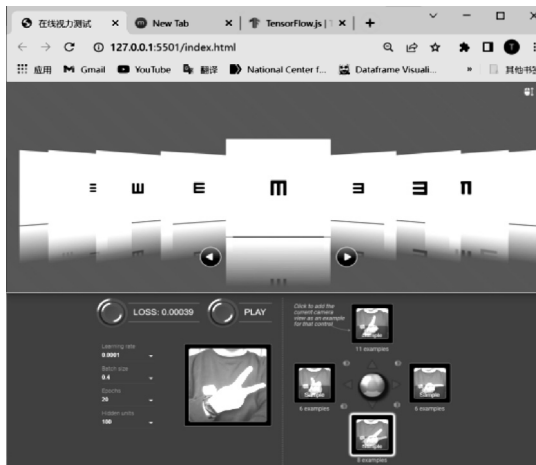


图 12 系统运行示意图 2

判断完一个视标后, 测试页面以轮播图的效果跳转下一个视标, 以此类推最后实现了视力的互动测试效果, 如图 13 所示。

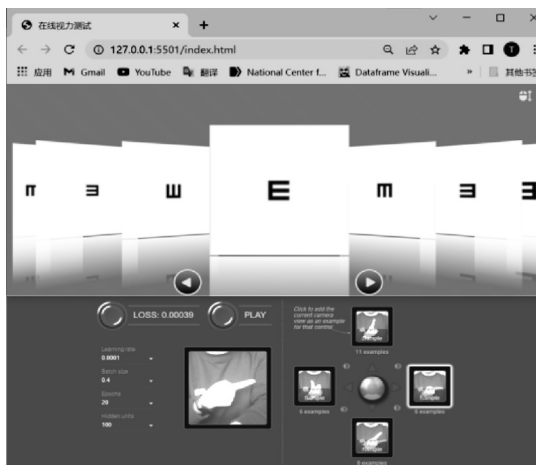


图 13 系统运行示意图 3

最后显示视力测试值为所能辨认的最小视标的视力记录值, 如图 14 所示, 左侧上部显示视力测试值结果为 4.8(0.8)。



图 14 视力测试值显示效果

6 结束语

系统运用 TensorFlow.JS 实现前端识别模型训练与应用, 借助屏显设备用摄像头捕捉人物动作姿态匹配对应类别, 使计算机生成的虚拟环境与客观存在的真实环境共存于同一个 AR 系统中, 并可在界面呈现实时交互效果, 最终实现系统进行 AR 互动测试视力的过程。从感官和体验效果上给用户呈现出虚拟对象与真实环境融为一体的 AR 环境。

未来的相关工作可围绕以下方面进行完善: 在测试训练之前, 用户应校准屏幕大小; 然后输入与屏幕的距离, 测试过程中在 4 个方向随机显示视标等来进行更精准的视力测试。

参考文献:

- [1] 侯颖, 许威威. 增强现实技术综述[J]. 计算机测量与控制, 2017, 25(2): 1-7.
- [2] 吴帆, 张亮. 增强现实技术发展及应用综述[J]. 电脑知识与技术, 2012(12): 8319-8325.
- [3] 朱森良, 姚远, 蒋云良. 增强现实综述[J]. 中国图象图形学报, 2004, 9(7): 767-774.
- [4] 替瑛瑛, 崔阿悦. 浅谈增强现实的现状与发展[J]. 科教文汇, 2019(13): 80-81.
- [5] 明德烈, 柳健, 田金文. 增强现实技术及其应用[J]. 无线电工程, 2001, 31(11): 51-54, 60.
- [6] QIN Z, ZHANG Z, CHEN X. Fd-MobileNet: Improved MobileNet with a Fast Downsampling Strategy[C]//2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 1363-1367.