

慧眼找物开题报告

一.应用背景

由于桌面杂乱、无收纳整理习惯等原因，在日常生活中经常存在一时无法从桌面上的众多物品中找到目标物品的情况。本App希望能够在该情况发生时，通过AR技术在桌面上标记物品的可能位置;若未能找到，则对物品所处的位置进行预测，以帮助人们快速寻找目标物品，节省时间。

二.用户分析

考虑到几乎全年龄段的人群均可能遇到难以在桌面上找出物品的情况，本产品希望能给全年龄段人群提供桌面寻物的快速寻找方案。但由于在对项目进行测试时，小组较难获取到青年以外人群的大量桌面数据，因此现阶段主要针对青年人群进行产品设计和测试。

三.设计方案



本项目预计将基于Android平台，完成一款慧眼找物手机App。接下来，就原型中交互设计，按照从上到下，从左到右的顺序逐一进行介绍。在上述画面中所有交互介绍完毕后，会对一些难以在原型中体现的交互设计进行介绍。

- 文字搜索

当用户不方便进行语音输入时，可在检索框中输入对目标物品的描述文本，App将基于该文本输入寻找物品

- 监控画面

本项目需配备一固定位置的外置摄像头，对桌面进行扫描，以便跟踪记录桌面上各物体的位置。在手机端App显示监控摄像头画面

- 输出提示

本项目提供语音助手，将语音作为输出，通知用户已在图中找到物品或预测目标物品位置。但考虑到部分用户不方便使用语音，会同步将语音提示信息在输出提示框中进行输出，方便用户读取。同时，与操作相关的部分提示信息也会在该提示框进行输出。

- 录入物品列表

系统支持用户录入物品，录入时需要提供物品的高清完整照片及自定义的物品名称。

如果用户希望寻找的物品在当前已录入的物品列表中，则将该物品与当前输入画面进行比对，并圈出目标物体。

- 语音输入

长按语音按钮，进行语音输入。App将对输入进行处理，并提取出用户希望寻找的物品信息，并在输入画面中对类似物品进行检索。

- 单机存储

在找到物品后，用户可以单击物品的正确位置进行物品信息存储，存储后的物品进入物品列表中，方便用户下次寻找。

四.创新点

在原有功能的基础上，小组成员经过讨论得出以下几个具有实现价值的创新功能。

- 如果物体存在部分的遮挡关系，系统可试图还原原物体轮廓
- 通过用户动作意图识别，预测用户要找的物品
- 用户在锁屏状态下通过语音告诉软件已经找到物品

上述创新功能中涉及目前本小组成员能力范围以外的技术，考虑到课程的时间限制，无法保证在项目最终答辩之前完成。

五.技术方案

- 外部接口

- 图像处理

- 百度智能云物体检测接口
 - 科大讯飞物体检测接口

- 语音处理

- 百度语言处理基础技术模块接口，包括词向量表示、词义相似度及短文本相似度等
 - 科大讯飞离线命令词识别

- 算法描述

- 图像处理

- 物体追踪 Multiple Object Tracking
 - 工作流程
 - 给定视频的原始帧
 - 运行对象检测器，获得对象的边界框
 - 对于每个检测到的物体，计算出不同的特征
 - 相似度计算，计算两个对象属于同一目标的概率
 - 关联，为每个对象分配数字ID
 - 参考方法

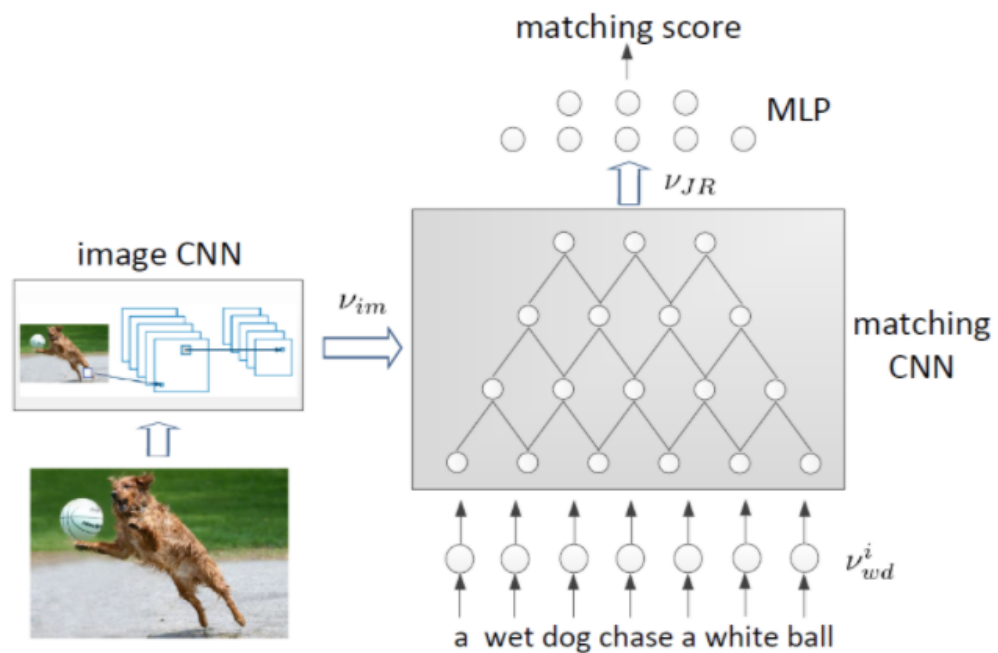
**** Multi-Object Tracking MOT Evaluation				
Methods	开源	来源	创新点概要	数据集
Real-Time Multiple People Tracking With Deeply Learned Candidate Selection And Person Re-ID (MOTDT)	是	ICME 2018	检测和跟踪结果作为候选对象，并基于R-FCN网络选择最佳候选者，轨迹评分机制	MOT15~17
Towards Real-Time Multi-Object Tracking (JDE)	是	Arxiv 2019	将检测器和嵌入模型集成到单个网络中 Triplet loss 自动学习损失权重方案	MOT15~17
Tracking without bells and whistles (Tracktor)	是	ICCV 2019	利用物体检测器的包围框回归来预测物体在下一帧中的位置，从而将检测器转换为Tracktor	MOT15~17
FAMNet: Joint Learning of Feature, Affinity and Multi-dimensional Assignment for Online Multiple Object Tracking	否	ICCV 2019 Oral	单个网络中完善了特征提取，相似度估计和多维分配 单目标跟踪技术和专用目标管理方案	MOT15~17、KITTI-Car
Spatial-Temporal Relation Networks for Multi-Object Tracking	否	ICCV 2019	时空关系网 (STRN)	MOT15~17
Robust Multi-Modality Multi-Object Tracking	是	ICCV 2019	传感器多模态融合模块 对点云的深度表示进行编码	KITTI-Car

知乎 @Harlek

■ 图像文本匹配

在得到各个主体位置后，可以将对应部分的图像与用户输入CNN网络进行图像文本匹配，用输出的匹配度得分衡量该图像与目标主体的相似度。

下图是使用CNN网络进行图像文本匹配，输出匹配度得分的具体过程。



○ 语音处理和语音助手

■ 语音命令

使用讯飞离线命令词识别接口，用户能够通过语音命令开启语音交互

■ 语音转文本

使用讯飞语音听写接口，将语音描述转成文字，实时返回结果，达到边说边返回的效果


○ 需要解决的问题

- 现有的文章和数据集多是针对行人、赛车和动物，怎样针对桌面物品做特征提取和跟踪？
- 怎样将语音输入与物品ID进行绑定？
- 目前希望将摄像机12h内的影像进行存储，并作为输入，长时间跟踪会不会导致寻找物品时间过长？能不能避免？
- 有没有可能将多目标追踪转成单目标追踪？
- 实时跟踪还是离线跟踪？

六.工作计划及分工

Week	Izumi	Edward	Naomi	谢安琪
4	学习图像处理		学习语音转文本	
5	完成物体检测和图像文本匹配demo		完成语音转文本demo	搭建前后端
6	整合demo与前后端，完成软件的技术原型迭代			
7				
8	图像处理其他基础功能实现整合		语音输入其他基础功能实现整合	
9				
10				
11	软件基础功能整体测试			
12	附加功能的相关技术学习			
13	附加功能的技术实现			
14	附加功能整合及测试			
15				
16	项目完善及答辩准备			

参考网址

- [1] 安卓语音助手 <https://github.com/just-ai/aimybox-android-assistant>
- [2] python 语音助手 <https://github.com/rcbyron/hey-athena-client>
- [3] 安卓 AR developers.google.com/ar/develop/java/quickstart
- [4] 物体检测 https://tensorflow.google.cn/lite/models/object_detection/overview
- [5] 物体识别，离线命令词识别 <https://www.xfyun.cn/services/object-recg>
- [6] 图像文本匹配 <https://www.csdn.net/article/2015-12-15/2826486>
- [7] Multimodal Convolutional Neural Networks for Matching Image and Sentence <http://mcnn.nohlab.com.hk/project.html>
- [8] tensorflow speech command https://github.com/tensorflow/tensorflow/tree/master/tensorflow/examples/speech_commands
- [9] 语音指令 <https://blog.csdn.net/rainhurt/article/details/89097842>
- [10] 多目标跟踪  https://zhuanlan.zhihu.com/p/97449724?from_voters_page=true
- [11] 单目标跟踪 <https://blog.csdn.net/jjydsb/article/details/86257635>
- [12] 单目标跟踪 (sort) <https://github.com/abewley/sort>
- [13] 单目标跟踪 (deepsort) https://github.com/nwojke/deep_sort

[14] 多目标跟踪综述 <https://blog.csdn.net/yuhq3/article/details/78742658>