|  |  |  |
| --- | --- | --- |
|  |  |  |

学堂在线线上监考考试 (Online Proctoring)说明书

|  |  |  |
| --- | --- | --- |
| **Prepared by:** 郑嘉文 | 13240946967 | **zhengyu@xuetangx.com** |

***仅限于学堂在线（慕华网络科技公司）内部使用***

*本文档所包含内容属于学堂在线（慕华网络科技公司）资产，本文档不能被用于，被复制，或者泄露给其他第三方，除非获得所有者书面授权。本文档的接收者，在使用本文档和保存本文档时段内，均同意保护以上所有者权利，防止丢失，失窃以及未经授权的使用*

版本历史

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本 | 时间 | 描述 | 修改人 | 授权人 |
| 0.1 | 2017年5月4号 | 创建说明书草案 | 郑嘉文 |  |
| 0.2 | 2017年6月1号 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# 参考文档

|  |  |  |
| --- | --- | --- |
| **版本** | **时间** | **文档名** |
|  |  |  |
|  |  |  |
|  |  |  |

02-12-99 9:11 AMTable of Contents

[参考文档 3](#_Toc482031111)

[1.0 问题定义 5](#_Toc482031112)

[1.1 问题或机会（Problem or Opportunity） 5](#_Toc482031113)

[1.2 目标 （Objective） 5](#_Toc482031114)

[2.0 市场分析（Market Analysis） 6](#_Toc482031115)

[2.1 现有的线上监考提供商 6](#_Toc482031116)

[2.2 共同点 6](#_Toc482031117)

[2.2.1 考试前 6](#_Toc482031118)

[2.2.2 考试地点 6](#_Toc482031119)

[2.3 不同点 7](#_Toc482031120)

[3.0 硬件要求 8](#_Toc482031121)

[3.1 摄像头 8](#_Toc482031122)

[4.0 软件 9](#_Toc482031123)

[4.1 人脸识别算法 9](#_Toc482031124)

[4.2 异物探测算法 9](#_Toc482031125)

[4.3 声音识别 9](#_Toc482031126)

[附录A: 前提和依赖（Assumption and Dependencies） 10](#_Toc482031127)

[a) 前提(Assumption) 10](#_Toc482031128)

[b) 依赖(Dependency) 10](#_Toc482031129)

[附录 B: 词汇表 11](#_Toc482031130)

# 问题定义

## 问题或机会（Problem or Opportunity）

随着慕课的风行一时，线上考试也流行起来。但是线上考试和传统的线下考试还是有很大不同的。线下考试可以统一组织，由老师监考。整套考试制度行之有效，非常成熟。与之相反，线上考试则是一个百分之百的新生事物。它给考生学习考试带来了很大的方便：

* Any time, Anywhere

考生可以选择想要的时间和地点考试

* 没有监考老师，或者至少考试现场没有监考老师（远程监考）

利害相生。线上考试给考生和学校带来便利的同时，也招致了很多质疑

* 如何确保线上考试的学生没有作弊
* 如何确保线上考试的学生没有串通作弊

重中之重就是如何保证考试结果的公正性 （Integrity）

## 目标 （Objective）

利用现有的硬件和软件资源，实现大规模的线上监考系统。主要考察

* 考察目前市场上现有的线上监考系统，分析其中的需求和不足
* 考察市场上现有的硬件，设备，机器价格
* 考察目前可以运用到线上监考系统的模型以及可能用到的算法
  + - * 机器学习（Machine Learning）
      * 深度学习（Deep Learning）

pinhole surveillance camera

- cheating - easily: print a fake ID, use a $20 KVM or just put a friend under your desk and use your laptop camera facing slightly upwards.

heard of KVM? You wont know I am suddenly controlling another computer (that can be as small as raspberry pie that is attached to the back of my monitor. In addition I can hide a HD camera somewhere in the back, these are tiny, and can record the whole session in HD.

What would be real evolution is actual virtual machines in exams where for instance you have to write a SQL code or similar. Not these silly drag and drop to fields "tests". I've done Cisco exams 5 years ago that had Router simulations...

A cheater won´t have to devote all their time to game the system. Kerri Davies told us how the security system works. Nothing special. Reto Egeter also mentioned a way to do it. I won´t post my ideas in here, but it´s simple.

# 市场分析（Market Analysis）

## 现有的线上监考提供商

|  |  |  |  |
| --- | --- | --- | --- |
| 产品 | 提供商 | 简介 | 网站URL |
| Online proctored (OP) exams | Pearson | 老牌的考试公司 | <http://www.pearsonvue.com/microsoft/op/index.asp> |
| MsLearn | Microsoft | 知名的软件提供商 | <https://borntolearn.mslearn.net/b/weblog/posts/online-proctoring-so-i-can-cheat-right> |
| eProctoring | Smowltech | 该公司的自动人脸识别技术是卖点 | <http://smowltech.com/en/our-service>  <http://www.smowl.net/> |
| ProctorTrack | Verificient |  | <https://www.verificient.com/> |
| RPnow | Software Secure | 为EDX提供监考服务 | <http://www.softwaresecure.com/>  <https://openedx.atlassian.net/wiki/display/EDUCATOR/Exams%3A+Proctored+Exams> |

## 公正性 （Integrity）

考试不仅仅是考生检验自身学习成果的工具，在现实中，考试结果成为了获取某种资源和资格的依据。所以考试的公正性是所有人都关心的最优先的问题。线上监考考试的公正性，招致了最大的质疑

## 共同点

### 考试前

* 提前检查计算机硬件配置，以符合标准
* 提前安装考试软件，验证能够运行样例考题
* 摄像头能清晰的看清楚考试地点周围的情况

### 考试地点

对地点环境的要求不完全一致，但大部分是一致的。都要求：

* 在封闭（四面围墙的）房子中，不能有其他人

如果考试过程中有人出入，考试成绩即作废

* 不能有其他噪音
* 不允许中间休息，上厕所也不行
* 在监考员可视范围/考生触手可及的范围内，不能有纸张（草稿纸也不行），手机，笔记，笔等物件
* 不能有双屏
* 不能有食物
* 清空口袋，取下手表，露出耳朵，并打开夹克衫
* 以电话，VOIP，internet与监考人员联系
* 身份ID必须一直可见
* 考试视频会被保存

## 不同点

主要的不同点是Smowl提供了人脸识别来辨别身份，而其他的系统仅仅识别身份证ID。Smowl 因为他们的人脸识别算法获得了Gartner的2014年度的Cool Vendor大奖

Smowl系统会在考试开始，考试结束时给考生拍照，并作为参数输入Smowl专有的生物识别模型。同时在开始进行过程中，Smowl系统也会随机选择机会拍照，并输入其识别模型，以确保考试过程中间没有换人。

# 技术难点及挑战（Challenge）

## 硬件设备

线上监考系统识别身份ID，一般有两种办法： 证件，和生物特征（比如指纹，人脸等）。但是无论哪种办法，系统都需要摄像。另外在监考过程中，系统也需要随机的照相，不间断的录像，以及在考试结束后向考试中心回传录像。为了更好更快更准确的识别身份，观察考生的考场行为，系统需要高精度，多功能的成像设备。而这样的设备大多价格昂贵。于此同时，线上监考系统需要录下考试现场所有的声音。

与之相对的是，价格昂贵的监考设备会对线上监考系统的推广造成困难。

## 单人作弊

单人作弊是指有考生自己策划并实施的作弊行为。比如打小抄，夹带等等。

## 串通作弊

### 替考

替考是指由他人代替考生本人完成考试。而考试的结果被认为是考生本人的成果。

### 远程作弊

远程是指考生利用现代化的通行设备或者软件（比如Messenger），向位于考场之外的个人或者团体寻求帮助，以提高考试成绩。

## 套题作弊 （Brain Dump）

还有一类考生，考试不以取得好成绩为目的，而是意图获得考试所用题库，从中牟利。他们使用的作弊工具有针孔摄像机，纽扣摄像机等等

# 线上监考模型研究



## 考前准备阶段

考生在报名考试以后，就进入考前准备阶段。在此阶段， 需准备：

* 自己的带相片的身份ID
* 线上考试用的电脑
* 领取线上监考设备

在考试当日，考试开始之前，一定要确定：

* 考试用电脑已安装考试软件，且一切工作正常
* 监考设备已调试完毕。光照，网路等满足要求
* 以某种方式和考试中心相联，比如电话，微信。。。

考生在做完一下动作后，进入5.2-- 考试阶段

* 考生照3张正面，侧面照
* 考生录完一段话上传
* 考生上传指纹
* 考生点击开始考试按钮

考试的规则，同市场上同类的软件，见2.2节

## 考试阶段

在考试阶段，考试系统会

* 随机的拍照，实时的识别
* 实时上传图像和视频
* 声音探测
* 识别报警

在此阶段，考生和监考老师以某种方式建立联系： 比如微信，电话，聊天工具等等。

## 考试结束

考试结束后， 考生需

* 再照3张正面侧面照
* 等待考试视频上传结束

## 事后检查阶段

考试结束后，考试组织者需事后审查考试视频。

* 由机器检查所有上传的考试视频，发现有问题的视频
* 由人再次复查有问题的视频，确认考试成绩的有效与否

## 图像识别算法

随着计算机算力的大幅度提高和数据的爆炸式增长，人工智能科学获得了突破性的进展。而其中机器学习和深度学习是当之无愧的明星。图像识别由于应用了神经元网络而效果大幅度上升，并已经走入实际的应用领域。公开的AI引擎也有很多，比如Google的tensorflow，百度的AI框架（Apollo？！#）

### 人脸识别算法

人脸识别是计算机视觉研究领域的一个热点。目前，在实验室环境下，许多人脸识别已经赶上（超过）人工识别精度（准确率：0.9427~0.9920），比如face++,DeepID3，FaceNet等。

但是，由于光线，角度，表情，年龄等多种因素，导致人脸识别技术无法在现实生活中广泛应用

#### MTCNN

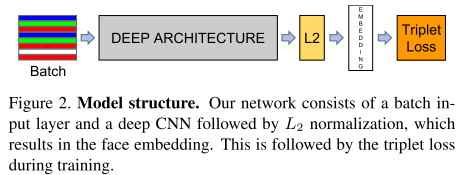
Multi-task CNN（MTCNN）



1. 采用opencv2实现从摄像头读取视频帧；
2. 对读取的视频帧采用mtcnn方法，检测人脸；
3. 采用预训练的facenet对检测的人脸进行embedding，embedding成128维度的特征；
4. 对人脸embedding特征采用knn进行分类，实现人脸识别；
5. 结果与改进；

#### Google人脸识别系统Facenet

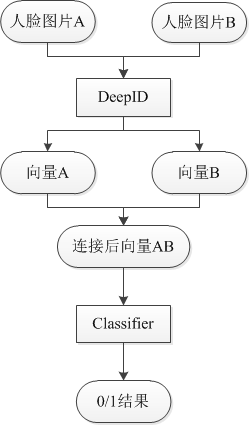
FaceNet是一个通用的系统，可以用于人脸验证（是否是同一人？），识别（这个人是谁？）和聚类（寻找类似的人？）。FaceNet采用的方法是通过卷积神经网络学习将图像映射到欧几里得空间。空间距离直接和图片相似度相关：同一个人的不同图像在空间距离很小，不同人的图像在空间中有较大的距离。只要该映射确定下来，相关的人脸识别任务就变得很简单。

当前存在的基于深度神经网络的人脸识别模型使用了分类层（classification layer）：中间层为人脸图像的向量映射，然后以分类层作为输出层。这类方法的弊端是不直接和效率低。  
  
与当前方法不同，FaceNet直接使用基于triplets的LMNN（最大边界近邻分类）的loss函数训练神经网络，网络直接输出为128维度的向量空间。我们选取的triplets（三联子）包含两个匹配脸部缩略图和一个非匹配的脸部缩略图，loss函数目标是通过距离边界区分正负类，如图1-1所示。

FaceNet是google的工作，工作量非常大，结果也很好。FaceNet是一种直接将人脸图像embedding进入欧几里得空间的方法。该模型的优点是只需要对图片进行很少量的处理（只需要裁剪脸部区域，而不需要额外预处理，比如3d对齐等），即可作为模型输入。同时，该模型在数据集上准确率非常高。

#### DeepID

DeepID达到的效果都是在LFW数据集上，该数据集是wild人脸数据集，即没有经过对其的人脸，背景变化比较大。该数据集太小，很多identities都只有一张人脸，5000个人只有13000张图片。所以DeepID引入了外部数据集CelebFaces和CelebFaces+，每次模型更新都会使用更大的数据集



在上述的流程中，DeepID可以换为Hog，LBP等传统特征提取算法。Classifier可以是SVM，Joint Bayes，LR，NN等任意的machine learning分类算法。

DeepID网络结构是香港中文大学的Sun Yi开发出来用来学习人脸特征的卷积神经网络。每张输入的人脸被表示为160维的向量，学习到的向量经过其他模型进行分类，在人脸验证试验上得到了97.45%的正确率，更进一步的，原作者改进了CNN，又得到了99.15%的正确率。

#### 其他

还可以用KPCA和SVM来进行人脸识别。

### 图像识别库

#### [OpenCV](http://opencv.org/)

OpenCV (Open Source Computer Vision Library) 支持计算机视觉和机器学习中很多算法。OpenCV自带C ++、C、Python、Java和MATLAB接口，并支持Windows、Linux、Android版和Mac OS等系统；

<http://opencvlibrary.svn.sourceforge.net/viewvc/opencvlibrary/trunk/opencv/data/haarcascades/>  可以下载各种分类器的haar特征数据。

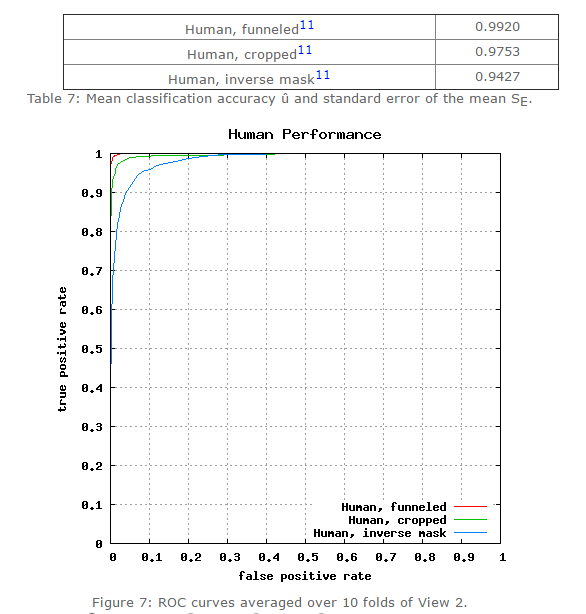
#### [DLib](http://dlib.net/imaging.html)

Dlib是一个用来处理图像管理和出阿里的C++库。DLib有C ++和Python接口，用于人脸检测和训练通用的目标探测器；

### 图像识别数据集

#### LFW

LFW数据集（[Labeled Faces in the Wild](http://link.zhihu.com/?target=http%3A//vis-www.cs.umass.edu/lfw/)）是目前用得最多的人脸图像数据库。该数据库共13，233幅图像，其中5749个人，其中1680人有两幅及以上的图像，4069人只有一幅图像。图像为250\*250大小的JPEG格式。绝大多数为彩色图，少数为灰度图。该数据库采集的是自然条件下人脸图片，目的是提高自然条件下人脸识别的精度。

目前，人工在该数据集上的准确率在0.9427~0.9920。在该数据集的第六种评价标准下（无限制，可以使用外部标注的数据），许多方法已经赶上（超过）人工识别精度，比如face++,DeepID3，FaceNet等。  
  
  


人类在LFW数据集上的识别精度

### ImageNet

Imagenet是一个图像数据库。官方网址是<http://www.image-net.org/about-overview> 。对于Wordnet里每一个有意义的概念（一个或者多个单词，又或者词组），Imagenet平均提供1000副图像。Imagenet对上面的图像并不拥有版权。

大多数的图像识别算法都在数据库[ImageNet](http://image-net.org/index)上进行人脸识别和分类用来做人脸识别。

### FaceTracer

FaceTracer数据库拥有巨大的真实人脸数据。FaceTracer还对FaceTracer上的数据集提供不同的属性标记和基准点。每一个子集合还有手工编辑的描述性属性，比如年龄，种族，胡子，头发颜色等等。总共有5000个属性。

网址： <http://www.cs.columbia.edu/CAVE/databases/facetracer/>

<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>

### 其他

可能使用的传统的和深度学习的算法有：

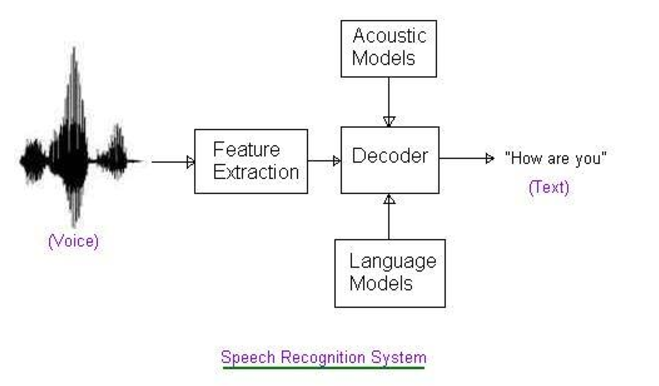
* R-CNN and Fast R-CNN
* Adaboost.

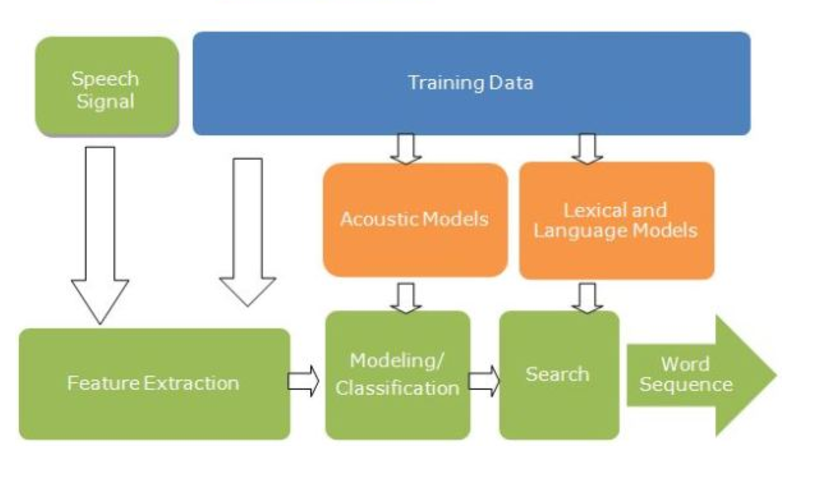
自适应地挑选分类精度高的弱分类器，将几个条件叠加，组成强分类

* 主成分分析（PCA）：一种线性降维方法，可以找出包含信息量较高的特征主成分，可以解释数据中的大多数方差。

## 声音识别算法

声音识别的算法主要基于隐含的马尔可夫模型（HMM），最大熵模型（Max Entropy），通过提取声音的特征值，来区分不同的人/动物/玩具的声音。在考试的环境下，主要的算法有两类： 声纹识别和声音探测。





### 声纹识别

声纹识别是为了判定当前说话的人和考生的声纹是一致，是身份认定的一种方式。基于的模型是“高斯混合模型”。不同说话人的差异主要表现在其短时语音谱的差异，而这又可以用每个说话人的短时谱特征所具有的概率密度函数来衡量。高斯混合模型GMM将空间分布的概率密度用多个高斯概率密度函数的加权和来拟合，可以平滑地逼近任意形状的概率密度函数，并且是一个易于处理的参数模型。在具体表示上，这个模型实际上就是把高斯混合模型的每个高斯分量的均值向量排列在一起组成一个超向量作为某一个说话人的模型，称为均值超矢量

可是，通常在实际中每一个说话人的语音数据很少，而训练高斯混合模型又需要大量的训练数据，怎么办呢？于是，UBM通用背景模型被提了出来。在训练说话人模型的时候，由于注册时说话人的数据稀疏，通常利用一个通用背景模型(Universal Background Model，UBM)和少量的说话人数据，通过自适应算法(如最大后验概率MAP，最大似然线性回归MLLR等)得到目标说话人模型。

用测试数据分别与模型和UBM进行似然度比较，然后将这两个似然相除再取对数，用得到的值作为打分来评价一条测试数据是否和模型匹配。

### 声音探测

声音探测是为了将一些奇异的声音从背景噪声中提取出来，作为考试行为异常的判定依据。此处的算法比较简单，并不识别奇异声音的语义并转化之，仅仅使用BP神经网络探测出某些不应该出现在考场的声音，并提醒考试系统注意而已。仅仅是一些波形和频率的匹配而已。

BP神经网络是ANN人工神经中的一种，常用的神经网络有BP、RBF、SOM、Hopfield等等，其功能不经相同，可总体来说ANN的主要功能是模式识别和分类训练

### 语音识别库

* [Kaldi](http://kaldi.sourceforge.net/)-Kaldi是用于语音识别的工具包，用C++编写，由Apache许可证V2.0协议授权，专门给语音识别的研究人员使用
* [CMU Sphinx](http://cmusphinx.sourceforge.net/) - CMU Sphinx 是基于Java 语音识别库，用于纯语音识别开源工具包。

## 运动视频识别

视频识别的大致使用如下的算法

1. 使用 ConvNet归类一帧
2. 从每一帧提取特征，将帧序列传给一个在独立网络中的RNN
3. 使用一个十分的ConvNet, 将帧序列传给一个在同一网络中的RNN
4. 从每一帧提取特征，将帧序列传给一个MLP
5. 使用3D卷积神经网络

# 线上监考系统设计

由于考试设备和考试的便利性是一对矛盾: 考试监视设备越多，则考试成本愈发昂贵，设备的安装调试也越复杂，考试也愈发不便；相对的，一味追求考试的便利性，虽然可以节省成本，但是却容易造成监视的疏漏，损害考试的公平性，进而危及考试的权威性。

所以，针对不同的重要性的考试，我们推出差异化的产品设计。

## 产品家族设计

### 简易考试

简易考试的着眼点在于尽可能的利用现有设备，降低成本。所针对的考试重要性不高。

设备：

* Laptop摄像头或者WebCam。
* 手机摄像头
* 电话，微信，QQ

### 严格考试

严格考试针对一些比较重要的考试，需要严格的监考的场合

* 360度无死角专用摄像头. 具体推荐设备见6.1
* 声音探测设备
* Raspberry设备（可选）, 具体推荐设备见6.2

### 集中考试

对于极端重要，极端敏感的考试，或者对于认为6.1.2监考方案过于繁琐，不方便的考生，建议采用传统的集中考试方法。

## 前提

使用线上考试的前提是

* 用户只能用一个屏幕
* 考试期间，用户必须呆在一个四面封闭的空间
* 考试期间，用户所处空间不能有任何人进入
* 考试期间，除了和考试中心外，用户不能和外界有任何形式的联系，电话，微信，email等。
* 考试期间，用户中间不能休息，离开房间
* 考试期间，用户不能进食

## 考试系统独占性设计

考试系统需具有以下功能：

* 一旦打开考试系统，考试系统将全屏显示
* 考试期间，用户不能用鼠标切换到其他程序。一旦发现考试切换，则认为考试未通过
* 考试期间，用户不能用快捷键，或者其他方式切换到其他程序。一旦发现考试切换，则认为考试未通过
* 考试期间，监考系统需监视考试用计算机，确认该计算机没有安装即时通信软件，比如Google Talk，微信，QQ等

## 考试系统静态头像识别

在上述的模型中，人脸识别，ID识别都需要用到图像识别算法。 图像识别主要用于识别考生身份以及分析考生在考场的行为，提取出有问题的场景，供监考老师复查，并适时报警。这其中有涉及到了

* 指纹识别
* 人脸识别
* 身体姿势识别（可选）

## 考试系统声音探测

为了防止考试现场混杂有其他人，动物，或者机器，在进行人脸识别的同时，系统必须要进行声音识别。以区别

* 环境噪音
* 考生自己的声音
* 其他人，动物，宠物，玩具，机器发出的声音

这其中有涉及到了

* 声纹识别以验证身份
* 声音探测仪探测异常

## 考试系统视频复查系统

尽管深度学习已经使得检测性能提升了一大截，但其实依旧存在许多难点。主要难点就是复杂光照情况（过暗、过曝）以及非刚性物体形变（如人体、手势的各种姿态）、低分辨率和模糊图片的检测场景。众所周知，目前大多数检测算法还是静态图的检测，而海量视频数据已然出现了，未来检测数据支持的类别肯定越来越多，涵盖的面越来越广，检测技术在这方面也需要继续发展。基于视频时序连续性的物体检测和像素级的实例检测将是未来重点突破的方向。

在线上监考系统中，在事后的视频复查中，我们也会做一定的尝试

## 考试系统树莓派专用系统

为了持续加强考试系统对各种考场环境的监视，系统提供一个接受服务器端指令的考试设备。该设备是一个树莓派设备：

* 能联网上传信息
* 能接受考试中心的各种指令，并返回运行结果

# 硬件要求

## 摄像设备

简而言之，线上监考系统需要一套摄像设备

* 能摄影
* 能录音
* 价格适中
* 能上网回传音频和视频数据
* 便携的

线上监考系统倾向于采购网络摄像机，360度无死角，最好带录音功能。

普通的电脑摄像头不适合做监控尤其是远距离监控。监控范围太小，且清晰度不够。另外USB线延长后，数据传输不稳定，长时间用数据丢失严重。而且电脑摄像头也经不起高的使用率。

网络摄像机本身集成有网卡，连上互联网就可以往考试中心回传数据。

由于摄像设备的采购和后面的人脸识别算法的要求密切相关，比如说精度，弧度（用来做图像处理的最好是平面图，而不能是变形的）。一旦图像识别的算法确定了，才好开始摄像设备的采购。

## Raspberry PI设备

# 参考索引

1. Pearson

<http://www.pearsonvue.com/microsoft/op/index.asp>

1. Microsfot MCSD/MCSE certificate

<https://borntolearn.mslearn.net/>

1. <http://smowl.net/>
2. <https://www.verificient.com/>

<https://www.proctortrack.com/>

1. <http://www.softwaresecure.com/product/remote-proctor-now/>

附录A: 前提和依赖（Assumption and Dependencies）

## 前提(Assumption)

|  |  |  |
| --- | --- | --- |
| **Assumption number** | **Description** | **Status (revised, remain, closed)** |
|  |  |  |

## 依赖(Dependency)

|  |  |  |
| --- | --- | --- |
| **Dependency number** | **Description** | **Status (revised, remain, closed)** |
|  |  |  |

附录 B: 词汇表

|  |  |  |
| --- | --- | --- |
| **简写** | **英文** | **中文** |
| CMS | Content Management System | 内容管理系统 |
| LMS | Learning Management System | 学习管理系统 |
| PDM | Product Data Mart | 产品数据目录 |
| WFS | Workflow System | 工作流系统 |
| UPP | User Profile and Privilege | 用户管理权限系统 |
| JSON | Javascript Object Notation | 一种轻量级的数据交换格式，基于ECMAScript的一个子集 |