# 人脸比对成功率情况说明

## 人脸比对应用场景分类

人脸比对（或称人脸识别）是指通过各种方法采集或抓拍截取了某人的人脸图像后，対对齐后的人脸图像提取其特征，与已知的某个数据库内的人员人脸特征逐个计算相似度，找到与查询特征最像的库内人员，最终通过固定或动态的阈值来确定查询的此人是否属于库内的人员，以及被认为匹配上的人员编号。通常按照被拍摄人是否配合、数据库是否固定等属性被划分为不同的应用场景，常见的应用场景如下表所示：

|  |  |  |
| --- | --- | --- |
|  | 需被拍摄者配合 | 被拍摄者无感 |
| 数据库固定 | 手机验证、交易验证、人证合一 | 公安人脸监控 |
| 数据库不固定 | 出入口闸机 | 案场新老顾客识别 |

被拍摄者是否配合是决定一个人脸比对系统成功率的非常重要的因素，在配合的场景下，人脸大小和方向都可控，难度要低很多，在无需配合的场景下，除了角度问题，还要面临分辨率低、遮挡等多种实际问题。下图显示了不同的应用场景下拍摄的人脸图片的范例（案场范例取自真实项目，其余取自公共数据集FRVT）：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **证件验证** | **手机验证** | **网络照片** | **某监控场景** | **案场场景** |
|  |  |  |  |  |

另一方面，数据库是否固定也决定了比对算法的根本差异。在固定库的场景下，通常被拍摄者可配合，因此系统对“不太像”的样本更倾向于拒绝匹配，要求被拍摄者再次尝试即可。而在非固定库的场景下，拒绝匹配的样本往往会被认为是新人而需要添加到库中，如果拒绝率太高则会导致库变得很碎，也会使人数等统计值变得不准。因此，对于非固定库的场景，调配拒绝匹配阈值变得非常重要。

## **影响人脸识别系统效果的因素**

一个完整的人脸识别系统由多个步骤组成，每一个步骤都会对最终的系统效果产生影响。常见的步骤有：成像、人脸检测跟踪、关键点检测、人脸质量评估、人脸特征比对等。所有前置步骤都会对后续步骤的效果产生影响。

### 成像

成像环节对最终的识别效果影响巨大，主要考虑的因素有：

* 人脸部分图像的大小：验证类场景下人脸图像宽高一般大于200像素，公安雪亮工程等专用场景下人脸一般大于100像素，而普通监控类场景人脸一般在70~100像素。人脸部分图像越小，对后续处理步骤提出的难度越大。
* 光照条件：是否拥有良好的光照条件使脸部区域更清晰。特别要注意的是室内向室外拍摄时，逆光场景下人脸部分会偏黑使细节缺失。

### 人脸检测跟踪

人脸检测跟踪模块负责从每帧图像中检测到人脸，并根据多帧的检测结果串联出人的行走轨迹，在视频中跟踪出每一个个体。这个模块决定了系统的抓拍率和重复抓拍率。主要考虑的因素有：

* 人脸的角度：检测器对人脸正面相对于相机的角度有一定要求，一般要求在横向纵向转动三个维度上都在30度以内，角度越正检测率越高。在无法要求被拍摄者配合的场景下，经常存在人低头、转头等情形，因此对检测器要求更高。
* 漏检与误检：场景中可能存在一些与人脸特征类似的物体，如影子、花盆、气球等，可能造成检测器误检，导致错误抓拍；同时某些人戴帽子、围巾等容易造成漏检。只有不断提高检测器的性能，让其在有限的算力条件下尽可能提高检测率、降低误检率，才能提高这部分的效果，满足现实使用需求。
* 人群遮挡：在无法要求被拍摄者配合的场景下，人之间的遮挡会较大程度影响跟踪的效果，导致一个人被抓拍多次、来向行人被错误跟踪到去向的人等等问题。

### **关键点检测**

检测到人脸后，由于人脸大小各异并存在角度偏差，需要检测人脸的额尖、双眼、鼻翼、嘴角等关键点位置，估计人脸在三个自由度上的角度，矫正角度并规范化图像。关键点定位的准确性决定了用来提取特征的图像的有效性。主要考虑的因素有：

* 面部遮挡：当面部有遮挡时（尤其是口罩），鼻翼、嘴角等位置定位失败，将影响对人脸大小及角度的判断，进而影响识别。这种情况下，一般的做法是拒绝对此人进行比对。在戴口罩较多的场合，也不得不退而求其次，在确认脸的上半部关键点定位较准确的前提下，按照上半部分的比例抠取人脸规范化图像（同时识别器也应使用戴口罩数据进行训练）。
* 旋转角度：人脸旋转角度过大时将导致关键点定位失败，一般要求在30度以内，而在非配合场景下，这一范围被放宽至45度乃至60度以内。

### **人脸质量评估**

抓拍过程的一个重要环节就是要选取某人行进过程中最好的一帧图像，最终用于提取人脸特征。其参考的信息包括人脸大小、角度、是否被遮挡、关键点质量等。一般人脸识别系统会设定一个宽容度阈值，当最终选择的图像综合评分小于此阈值时，将拒绝对此人进行比对。此综合评分各家定义不同，也很不直观，一般不向用户暴露此参数。

### **人脸特征比对**

人脸质量评分超过阈值后，将规范化图像输入深度学习网络提取人脸特征，并与库内的所有特征逐一进行比对，当相似度最高分超过一定阈值，即认为匹配成功。在实际系统中，用两个指标来描述比对模块：拒绝识别率（FNMR）和错误识别率（FMR）。拒绝识别率表示本应该认为是同一个人被算法认为不同的数量占全部测试数量的比例，错误识别率表示本是不同人被算法认为相同的比例，两个率是相互制约的关系。在比较比对算法时，会将错误识别率固定在一个定值，比较拒绝识别率，越低越好。

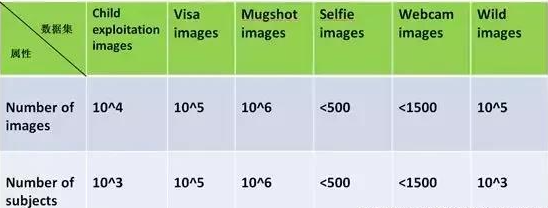
在不同的应用场景下，两类错误的代价往往不太一样，比如金融领域的人脸识别，误识会是一个很严重的事故，相比之下，拒识结果相对可以接受。而在案场新老顾客场景，则对误识相对宽容，对拒识相对要严格一些（不希望把老顾客认为是新顾客）。匹配效果主要考虑的因素有：

* 数据库规模：需要比对的人的ID数量越大，一个人被误匹配的概率就越高；相应的，在固定的误匹配率下，拒识率也就越高。在最差情况下，一次查询的正确率与数据库规模呈指数关系。也即，假设一个比对系统的1:1错误率在万分之一，即正确率为99.99%，那么如果用来在一个规模为1万个ID的库里进行一次逐一比对，其正确率在最差情况下将下降到0.999910000=36%。一般来说，针对人数为N的库，识别算法的1:1错误率应在1/10N以下，才能保证其实用性。因此，不能盲目对比单纯的图片比对错误率，而应该结合库的规模，在整体系统级别上衡量。
* 比对速度：包括提取特征的速度和比对速度。一般对于百万级大库比对的应用场景来说比对速度这个因素比较重要。而对于需要及时反馈的系统，更重要的是前端抓拍延迟、网络传输延迟，以及云端处理队列的优化。

## 当前行业实际水平

近年在各种新闻稿中，关于人脸识别模块的效果好到超过人类之类的新闻层出不穷。出现了很多公开测试数据集，例如LFW，IJB，MegaFace等等，这些数据集合的一大显著特性是从互联网上下载并且整理的一些图片，很难反应实际应用中的一些需求。

在这些公开数据集中，由NIST主持的FRVT测试数据集在一定程度上能反映实际应用下的效果。它是由多种测试集合组成（比如受控环境下人脸证件照片VISA、Mugshot，人脸自拍照片Border、一般互联网照片Wild、监控类Child Exp等，样例照片可参考上节表格）。从样例照片可以看出，案场应用的库规模及识别难度大约在Wild和Child Exp之间。各个测试集的图片数量和ID人数如下表所示：



在FRVT的测试报告中给出了特定的FMR指标阈值下FNMR的单项测试集合的结果和排名，即允许在某个固定比例的不同人分到同一人的情况下，会把同一个人认成不同人的比例。下表显示了国内公司目前在这个测试集上的1:1效果及排名：

|  |  |
| --- | --- |
|  |  |
| 格林深瞳 |  |
| 大华 |  |
| 商汤 |  |
| 旷视 |  |
| 滴滴 |  |
| 腾讯 |  |

从测试集的指标设计和各算法的效果上就能看出，不同场景的实际应用效果存在巨大的鸿沟。同样是高分辨率图像，配合类场景（VISA）比不配合类场景（Wild）的算法性能指标FMR和FNMR二者均相差了一个数量级，而Child Exp场景的性能指标与前述场景相差巨大，目前来看没有算法可以达到实用级别。

## 文安技术水平

鉴于文安在实际项目中碰到的场景与网络公开测试集的区别较大，文安并没有跟进参与FRVT测试，而是选择在实际项目中直接测试综合指标。针对上述各个步骤需要考虑的问题，文安相应的模块技术说明如下表所示：

|  |  |
| --- | --- |
| **分步骤考虑因素** | **文安解决方案及技术参数** |
| 光照条件（逆光） | 专为逆光环境设计了WDR模式，使各种光照条件下人脸成像均达到最好 |
| 人脸分辨率较低、人脸偏转角度大 | 人脸检测、关键点对准、识别模块均使用超过100个不同项目的真实数据训练，可适应较低分辨率和较大角度范围，支持最小人脸像素60\*70，横向角度±50度以内，纵向角度0~45度以内 |
| 漏检和误检 | 检测器模块的检测率>99%准确率>99%，并可提供模型进化服务快速适应新场景 |
| 人群遮挡 | 检测器可同时检测正脸、侧脸和后背并分类，给跟踪模块提供更多信息，减少行人间轨迹交替造成的跟踪错误 |
| 面部遮挡（口罩等） | 增加了口罩检测逻辑，对上半脸特征较明显的抓拍也提取特征用于比对，实际项目中约有30%~50%戴口罩的人可以提取到质量过关的特征用于比对 |
| 人脸质量过滤 | 可使用人脸质量过滤条件来去除由于人脸大小及人脸角度超过可识别范围、面部遮挡较大、手部遮挡、其他人遮挡、成像不佳等原因带来的不佳抓拍（不影响客流数，只是无人脸特征）。实际项目中被过滤的比例可达10%~30%不等。此过滤条件可控，可在系统试运行期间由文安技术人员统计并配置 |
| 比对成功率 | 对案场场景，不带口罩情况下，在FMR<0.00001（十万分之一，用于万人库）情况下，FNMR<0.1；在FMR<0.0001（万分之一，用于千人库）情况下，FNMR<0.06 |
| 提特征及比对速度 | 在私有云部署的环境下，如需后端提特征，平台可在1秒内完成提取特征操作，库容量在100w以内的比对操作可在100毫秒内完成。 |