目录（示例）

第一部分毕业论文

第二部分 文献综述和开题报告

1. 引言

自古以来，每个人都有着自己的身份，有代表君王身份的玉玺，代表将领身份的虎符，无一不是用来识别身份的。而就现代而言随着信息技术的进步，随着网络应用的推广，网络交易时代的到来，个人身份的识别无疑更为重要。近几年来，人脸识别技术逐渐成为业界研究热点，因为人脸识别与 DNA 识别、虹膜识别相比，有着良好的用户体验，有着便捷的操作。而随着人脸识别技术的逐渐研究，技术日益成熟，人脸识别技术的逐渐商业化也给人脸识别技术的一些安全性、稳定性、准确性提出了更高的要求。

目前的技术来说，人脸识别虽然简捷快速，可是也有这一些致命的缺陷。因为用户的生理特征并不严格保密，相比于指纹、虹膜等，人脸的照片，人说话的视频尤其易于获得，而一些人脸识别系统也可以用照片、视频等重放手段来破解，会给人脸识别系统的可靠性带来巨大的挑战，所以人脸识别系统中的活体检测成为了必须的手段。

人脸识别中的活体检测是指系统通过收集获取用户的一些行为特征，如笔迹、说话、姿态等或一些生理特征，如人脸、指纹、虹膜等[1]，来判断用户是否是活体，是否是本人。

1.1 研究背景与意义

活体检测概念源于生物识别技术领域，生物识别技术在上个世纪已经有了一定的发展，所谓生物识别技术就是，通过计算机与各种信息传感器和生物统计学原理等高科技手段密切结合，利用人体固有的生理特性，如指纹、人脸特征、瞳间距等和行为特征，如写字习惯、声音音调、走路姿态等来进行个人身份的鉴定[2]。其中指纹识别技术已经臻于完善，但是人脸识别技术的开发仍处于起步阶段。

指纹、虹膜、掌纹等识别技术都需要大量的人机交互，需要待识别用户的配合，甚至还需要昂贵的外加仪器，而人脸的活体检测可以在被识别者不知道的情况下进行，这对反欺骗系统具有非常重大的意义。传统的生物识别技术应用于考勤市场、门锁市场、门禁市场、重点行业内部人员身份认证市场等[3]，而人脸活体检测大多服务于智能监控领域、金融服务领域、商业服务领域等对安全性要求更高的领域。

当人脸识别系统逐渐开始商用之时，人脸活体检测成为人脸识别的关键组成

部分，它对于任何一个生物检测系统来说是十分重要和必要的一个环节，它可以保证生物检测系统能够安全有效地工作；对于无人监督的人脸识别系统应用来说，自动地抵抗照片和视频欺骗是一个人脸识别领域中一个迫切需要解决的问题。

1999 年 T.Choudhury[4]较早地在系统中设计了用三维深度信息区分照片和活体，而在 2002 年，L.Thallheim[5]等人在 c't magazine 一文中发表了用合法用户的人脸照片和视频可以成功欺骗系统，通过系统认证。正是因为这种人脸识别系统中存在的可能引发严重后果的漏洞引起了人们对于活体检测研究的热潮，活体检测技术得以快速发展。

在近十几年来，人脸识别技术的识别准确率不断提高，对环境的抗干扰能力不断提高，为人脸识别活体检测的发展做下了铺垫。目前为止，人脸识别活体检测已经开发出来七种左右的不同的，可以相互融合的反欺骗手段，为生物识别领域做出来巨大的贡献。

1.2 国内外研究现状

1.2.1 人脸识别活体检测的欺骗手段

对于可见光图像的人脸识别系统来说，最常见的欺骗手段有三种：使用合法用户的人脸照片（打印照片、手机屏幕照片）进行欺骗、使用合法用户的人脸视频（手机视频、电脑视频）进行欺骗、使用合法用户的三维人脸模型进行欺骗[6]。

照片是获取最为简单的欺骗工具，攻击者可以轻易地在合法用户不知情的情况下进行偷拍或是在合法用户的个人主页、博客等社交平台获取各种照片。不过照片缺乏动态信息，可以较为简单的进行预防。

相比之下，视频欺骗更具威胁性。攻击者仍旧可以通过使用摄像机偷拍、使用针孔摄像头偷拍等获取。与二维平面的照片相比，视频可以体现更多的信息，包括眨眼、点头等行为特征，预防起来相对更难一些。

目前为止，三维人脸模型的欺骗手段不太常见，攻击者可以通过使用多幅人脸图像，通过变形来建立三维人脸模型，并且可以合成多种表情来通过检测，因为三维人脸模型的数据相比于照片和视频更难获取，而且制作成本高昂。不过相比于前两者，三维人脸模型的威胁性最大。

1.2.2 已有的抗欺骗手段介绍

近年来，活体人脸的检测主要有以下几种线索：三维深度信息、脸部运动的光流信息、人脸和语音混合信息、傅里叶频谱信息、眨眼信息、人体红外信息等。下面将对这几种方法进行简单的介绍。

1.2.2.1 三维深度分析

真实的人脸是三维立体的结构，而照片与视频当中的人脸是二维结构，无论是移动、折叠还是扭曲，能够提供的三维信息都无法与真实的人脸相比。

三维深度分析的主要思想史通过 SfM（Structure from Motion）模型估计出人脸上若干特征点在三维坐标系中的深度坐标[7]。人脸上的眼睛、嘴巴和鼻子可以作为三维深度坐标的特征点，而视频照片里面的人脸是平面结构，照片上的特征点的三维深度坐标都十分接近。所以三维深度信息可以用以作区分真人与照片视频。

1.2.2.2 脸部运动的光流估计

光流对于物体运动比较敏感，Kollreider 等人提出了用光学流的方法检测人脸动态变化从而进行活体检测的办法[8]。光流法利用图像序列中各个像素点的强度数据的时域变化和相关性来获取各自像素的运动信息，采用高斯差分滤波器、LBP 特征和支持向量机等进行数据统计分析[9]。

对于照片来说，脸部不同的特征点的运动幅度、方向、速度是相同的，而 3D 活体人脸上不同的区域，靠近中间的区域（如鼻子耳朵）产生的 2D 运动大小相比于外部区域（如耳朵）要大。下图是头部从左向右转动时的水平方向光流情况。

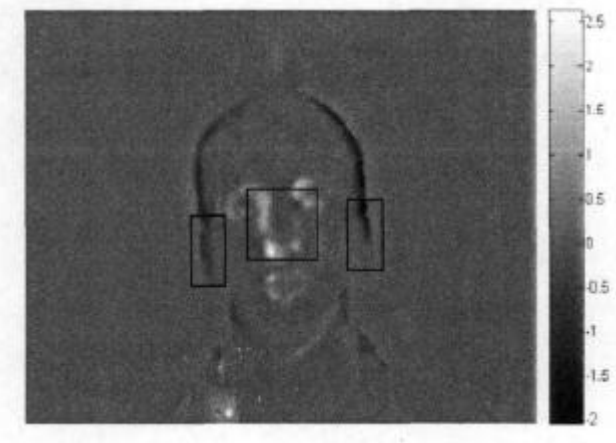

图 1.1 水平方向光流分布[8]

如下图所示，可以根据光流场中的数据来区分活体和照片。

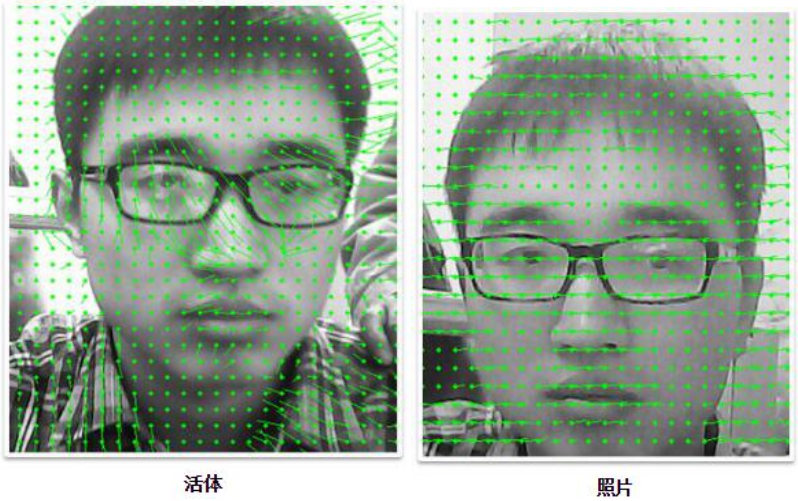

图 1.2 活体图像与照片的光流场分布[3]

不过根据 Kollreider 等人的实验表明，光流法可以有效地区分活体和照片，但没有办法抵抗视频欺骗，而且检验时的光线会严重干扰该方法的准确率。

1.2.2.3 人脸和语音混合模型

在单一人脸识别身份认证系统当中，存在照片视频欺骗的情况，而且会受到单一的人脸识别现有技术的识别能力的限制。因此可以采用人脸和语音混合模型的办法，甚至可以采用指纹、声音、人脸等多模生物特征融合的认证方法。

如果说单一生物特征容易获取，那么多种生物特征，特别是有一些特定生物特征，如按照认证系统指示说一句话，就要难得多，而且多模混合模型当中，每个生物特征都可以单独评分，欺骗起来十分繁琐。因此可以有效提高生物认证系统活体检测能力。

G.Chetty and M. Wagner[10]认为，在说话和相应的脸部运动时，声音动态特征直接存在着相关性，表现在嘴唇、舌头上。而人脸语音混合模型中，使用嘴唇位置的动态特征来进行活体检测是一种自然有效的方法，这个方法中有两个问题需要解决，一个是找出和跟踪嘴唇，一个是识别，也就是估计出与嘴唇布局相关的特征数据。下图是寻找嘴唇和定位嘴唇以及分析的例子。

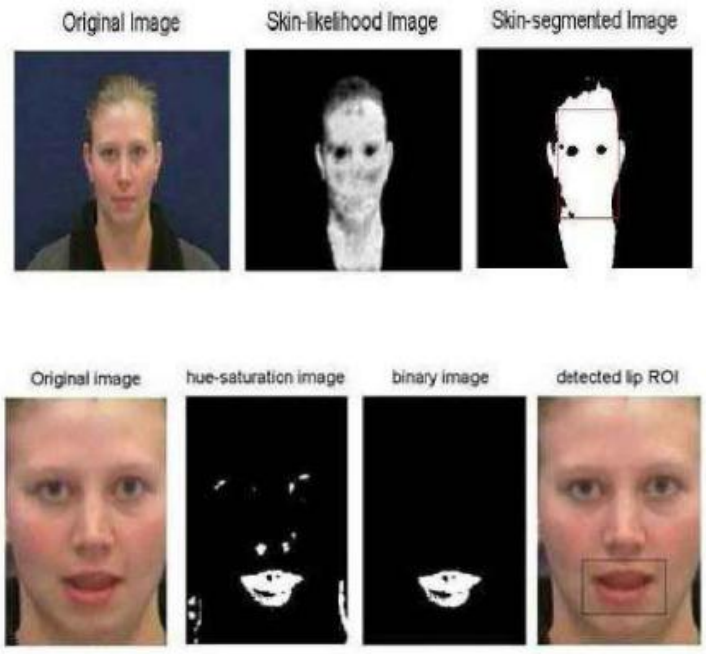

图 1.3 嘴唇分割及定位的例子[10]

1.2.2.4 傅里叶频谱分析

Jiangwei Li[11]等人认为，傅里叶频谱分析可以用来区分活体人脸和照片。Li 等人认为，照片是二维平面，照片的高频分量要相比活体人脸的成像少，而且因为活体人脸具有丰富的表情，所以活体人脸的频率分量的标准差应该要高于照片，下图展示了活体人脸和照片人脸在频域上的差异。

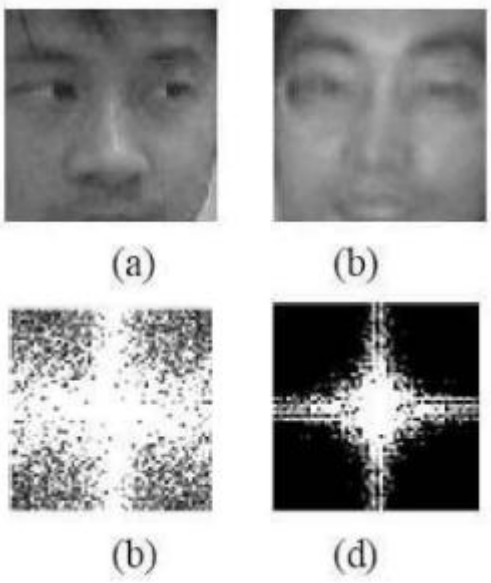

图 1.4 活体人脸和照片人脸在频域上的差异 [11]

（a）活体人脸（b）照片人脸（c）活体人脸频域图（d）照片频域图

相比于前几种方法，这种方法更加易于实现，但是对于照片欺骗的抵抗程度不高。

1.2.2.5 眨眼检测

眨眼检测在几种方法中是较为简单的一种，该方法可以将眼睛分为张开和闭拢两种状态，可以用一个 SVM 分类器来进行判断[12]。眨眼检测系统可以给被检测者连续拍几张照片，进行判断眼睛的张合情况，如果又有张开又有闭拢的状态，则判断为活体，只有一种状态的判定为照片。不过仅通过眼睛的张闭来判断是否为活体的缺点是无法有效抵抗视频欺骗。

1.2.2.6 热红成像

目前来看，热红成像是活体检测最为有效的一种方法，与可见光图像不同，热红成像直接展现了人脸热辐射的情况，而且有人脸热辐射本身就表示被检测者是活体，所以对于照片、视频欺骗的抵抗能力极强。下图是可见光照片与红外线光照片的对比。

图 1.5 可见光与红外线光照片对比[6]

但是，热红成像仪的价格也十分昂贵，是可将摄像头的上百倍，在成本受限的情况下，热红成像并不是最优选择。

1.2.3 已有方法的分析对比

几种不同的方法在用户配合程度、抵抗光线干扰能力、是否需要外加设备、照片视频欺骗抵抗能力上进行了对比，对比结果如下。

表 1.1 不同方法的对比表

方法	用户配合程度	受光线影响程度	附加设备	抵抗照片欺骗	抵抗视频欺骗
三维深度	一般	中	无	一般	弱
脸部运动的光流估计	少	中	无	较强	弱
人脸和语音混合	一般	低	有	/	/
傅里叶频谱	少	中	无	一般	一般
眨眼检测	少	低	无	较强	弱
热红成像	无	低	有	强	强

综合以上的比较，我认为最为理想的活体检测方法应该满足以下几点条件，以后的人脸活体检测工作应该会往这几个方面发展：

1、不需要用户主动配合；

2、不需要外加设备；

3、受外界因素干扰小；

4、对于照片视频等欺骗攻击的抵抗能力强；

5、总的计算量小，可以快速得出验证结果。

1.3 主要研究内容

1.3.1 主要研究内容

本课题旨在根据实验室已有的人脸识别技术、人脸区域提取技术、人脸特征点提取技术、卷积神经网络技术，研究开发具有良好抗欺骗能力、优秀抗干扰能力、仅需用户少量配合的、仅需普通摄像头（无需外加设备）、融合的人脸识别活体检测技术。

本研究主要分眨眼检测、摩尔纹检测、黑框检测三个部分，之后将三个检测融合得出最终的活体检测结果。

1.3.2 技术路线

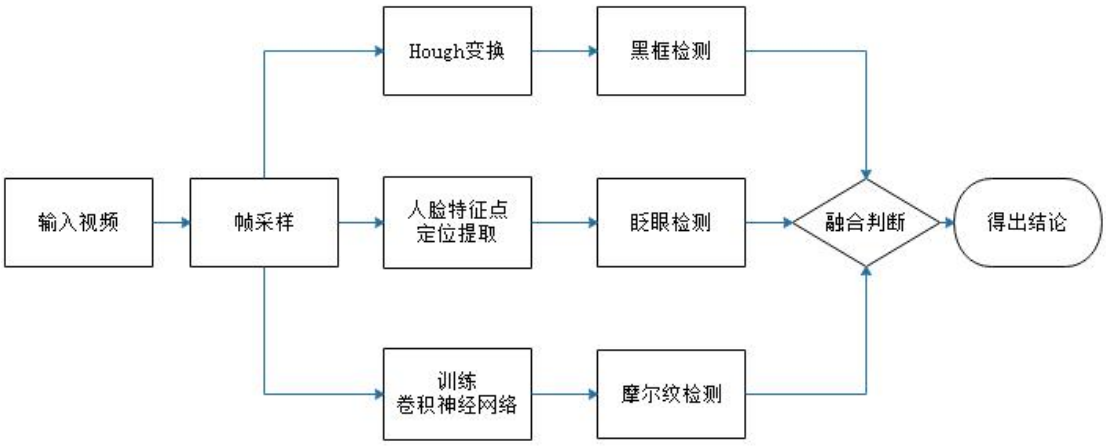

图 1.6 技术路线图

2.黑框检测的研究

2.1 问题描述

传统的人脸识别是通过摄像头或相机直接获取用户信息，来判断进行识别的用户是否为合法用户。非法用户通过获取合法用户的照片，并把照片置于摄像头前即可通过没有抗欺骗手段的人脸识别系统。尤其到了 20 世纪，随着互联网技术、移动设备技术的发展，人们的照片已经不再是隐私，从网上下载合法用户的照片、视频再以之欺骗人脸识别系统可以说是十分便捷。而在非法用户使用手机上的合法用户的照片、视频的时候，我们认为，或多或少会产生一些如下图的黑框。

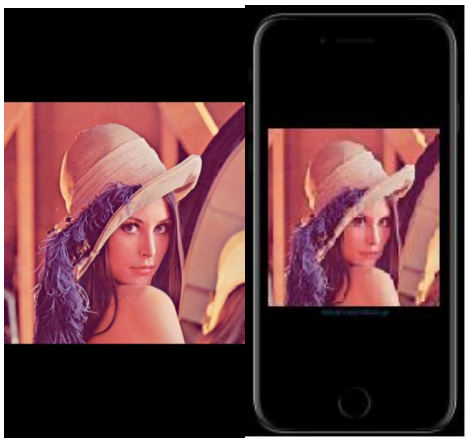

图 2.1 黑框图

如图所示，照片与上下的黑框差异十分显著，而且非法用户使用手机上的照片的时候，照片与手机屏幕之间或者与周围环境也会产生差异显著的黑框。因此通过黑框检测，可以快速有效地降低对于数字照片、视频的攻击的抵抗能力。

2.2 研究方法

2.2.1 canny 边缘提取

高斯滤波图像去噪：$f(x,y)$为输入数据，$G(x,y)$表示二维高斯函数，$f_x(x,y)$表示高斯卷积平滑之后的图像。其中：

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{\frac{-(x^2+y^2)}{2\sigma^2}}$$

$$f_s(x,y)=f(x,y)*G(x,y)$$

计算图像梯度：使用 Sobel 算子对$f_x(x,y)$图像计算水平和竖直方向上的图像梯度（G_x和G_y），根据G_x和G_y找到边界梯度及其方向。其中：

$$Gradient(G)=\sqrt{G_x{}^2+G_y{}^2}$$

$$Angle(\theta)=\tan^{-1}(\frac{G_x}{G_y})$$

非极大值抑制：将角度划分成0°、-45°、90°、$+45^\circ$，再对访问点的 3*3 领域内的四个方向进行非极大值抑制，即在沿其方向上领域的梯度幅值最大则保留，否则抑制。

双阈值检测：选取高阈值和低阈值，一般$T_H=0.3/0.2$，$T_L=0.1$，重新定义高低阈值，即$T_H\times\max, T_L\times\max$，将小于低阈值的点置 0，大于高阈值的点置 1，将介于两个阈值中间的点使用 8 连通区域确定。

2.2.2 hough 变换

根据直角坐标系点(x,y)与极坐标系点(ρ,θ)之间的关系$\rho=x\cos\theta+y\sin\theta$（$(\rho,\theta)$是坐标原点到直线的距离和垂线与 x 轴的夹角）可以得出结论，在直角

坐标系中的直线可以转化成极坐标系下的一个点。同理，对于每一个极坐标下的点，都存在一系列的曲线通过这个点，

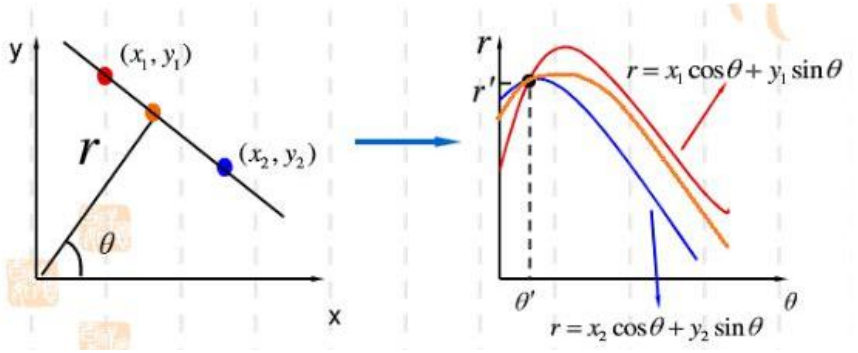

图 2.2 hough 变换原理图

图中蓝、橙、红色三条极坐标曲线同时经过一个点，而极坐标上每一个点对 (ρ,θ) 在空间坐标上对应一条直线。同时经过的这点表示在直角坐标系下，有一条直线经过直角坐标系下的红点，橙点，红点，说明这三个点可以构成一条直线。

2.2.3 提取边界直线

首先将图片重定义大小为 200*200 的图片，再对图片进行灰度化处理，使 canny 边缘提取算法提取灰度图的边缘，再将边缘图进行 hough 变换。

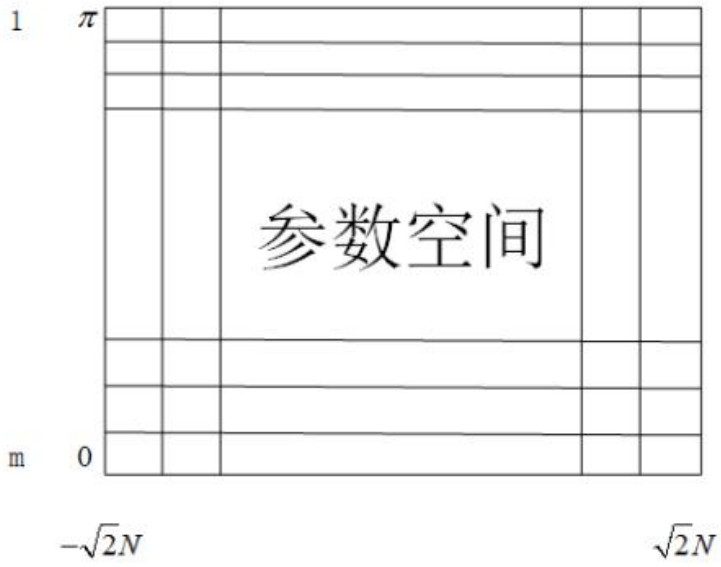

图 2.3 hough 变换提取边界直线原理图

将边缘图分成 200*200 个网格。

即将$\theta \in [0,\pi]$分成 200 份，将$\rho \in [-\sqrt{2}N,\sqrt{2}N]$分成 200 份，然后设定一个 200*200 的累加单元，用来存储图像中某一条直线出现的次数。对边缘图中每一个像素点(x,y)，根据公式$\rho = x\cos\theta + y\sin\theta$计算$(\rho,\theta)$，然后再相应的累加单元加 1。

之后在 hough 变换图中寻找 100 个像素以上的直线形成的点。统计直线检测结果中直线包含像素的个数，超过一定阈值（30）则判断为黑框。

2.3 结果验证和分析

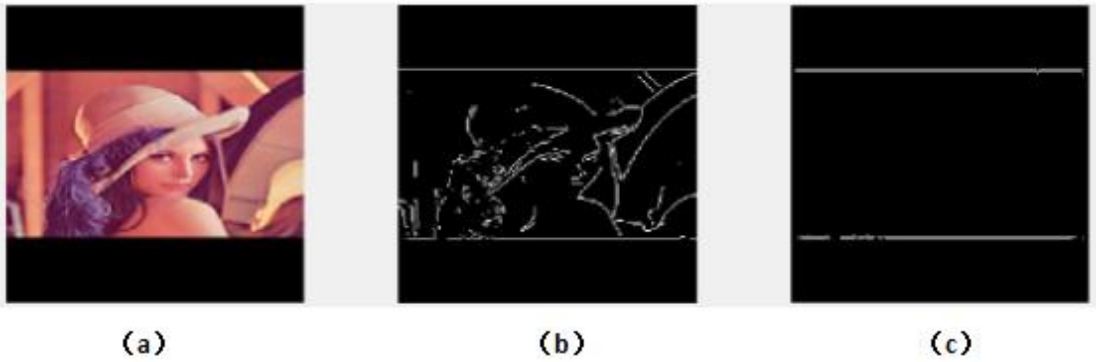

图 2.4 （a）原图（b）canny 边缘提取图（c）hough 变换提取边界图

如图所示，a 为原图，b 为使用 canny 算法提取的边缘图，c 为使用 hough 变换提取的边界图，可以看到，黑框提取的较为完整。

本课题使用实验室私有数据集进行测试，私有数据集包含共 142 个对象的不同黑框照片，使用本算法进行黑框检测，检测的准确率可以达到 97.2%。

不过在使用摄像头进行拍摄识别的时候准确率会有所降低，该算法受光线的影响比较大，良好的光线是本算法识别的基础。

3.摩尔纹检测的研究

3.1 问题描述

如下图，摩尔纹是一种在数码照相机或者扫描仪等设备上，感光元件出现的高频干扰的条纹，是一种会使图片出现彩色的高频率不规则的条纹[13]，并没有明显的形状规律。

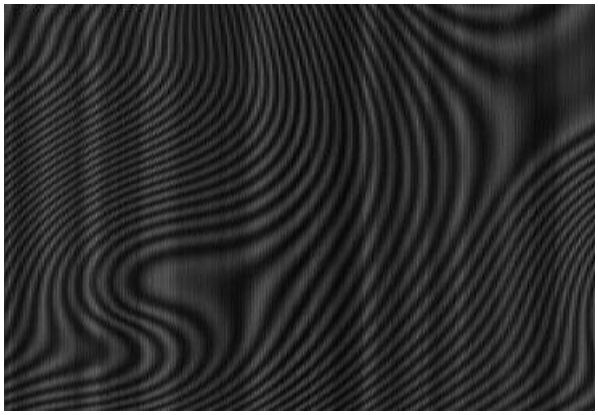

图 3.1 摩尔纹图

从目前的技术来说，在应用到人脸活体检测的场合中，人脸识别虽然简捷快速，可是也有这一些致命的缺陷。因为用户的生理特征并不严格保密，相比于指纹、虹膜等，人脸的照片，人说话的视频尤其易于获得，而一些人脸识别系统也可以用照片、视频等重放手段来破解，会给人脸识别系统的可靠性带来巨大的挑战。而通过直接拍摄照片、间接录制视频、用手机拍摄网上的照片等手段获取的合法用户的资料，容易带有摩尔纹。因此通过摩尔纹检测，可以有效加强人脸识别系统对于照片、视频重放攻击的抗欺骗能力。

3.2 研究方法

3.2.1 深度残差网络

深度网络的好处可以概括为两点：一、网络可以提取的特征所包含的信息量随着网络的深度的加深而变得更大；二、深度足够大的网络具有极其强大的表达能力。不过在大深度的网络训练当中，会遇到梯度弥散和退化问题两个问题[14]。

梯度弥散：在优化基于反向传播计算梯度的神经网络时，使用链式法则来求取隐藏层的梯度，所以会经过一系列的连乘来得出梯度值，导致浅层的隐藏层会出现剧烈的衰减。

退化问题：深度过深的平原网络的训练误差更高。

深度残差网络中的残差块可以有效解决这两个问题。

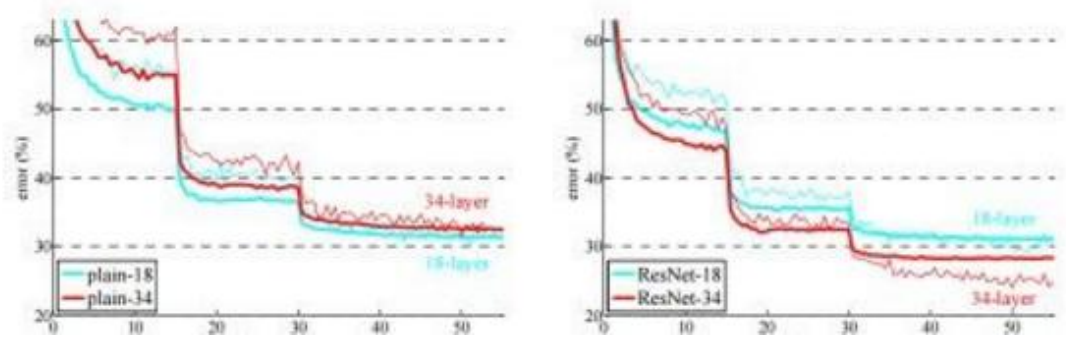

图 3.2 18 层和 34 层网络的平原网络和残差网络对比图

从上图可以看出，对于 Resnet 网络，较深的模型比较浅的模型有更低的训练误差。

残差块结构[15]如下：

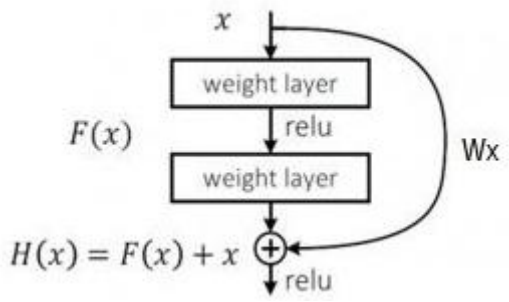

图 3.2 残差块图

残差块有二层，其中 $F = W_2 ReLU(W_1 x)$，之后通过一个 shortcut 和第二个 ReLU，获得输出 $y = F(x, \{W_i\}) + x$。残差块是的拟合对象变成 F（x），拟合目标变成是 F（x）趋近于 0。

如果输入 x 和输出 F（x）的维度不同，在 shortcut 上对 x 做一个线性变换 Ws 就可以解决这个问题，使得 $y = F(x, \{W_i\}) + W_s x$。

3.2.2 Resnet50 网络结构

本课题使用的 Resnet50 网络，属于深度残差网络，结构如下：

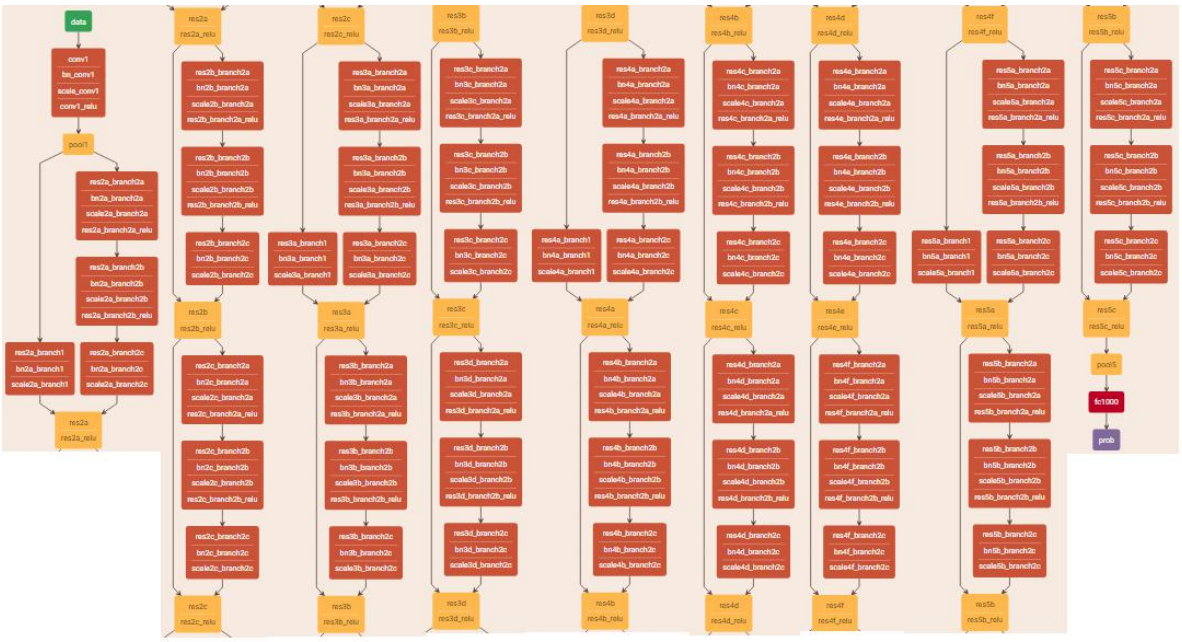

图 3.4 Resnet50 网络结构图

本网络的输入是 224*224*1 的灰度图，输出是 1*1*2 的特征值，整个网络可以分成 8 个部分：卷积层、最大池化层、resnet_2 层、resnet_3 层、resnet_3 层、resnet_4 层、resnet_5 层、均值池化层、全连接层。

其中，还包括特殊层：Batchnorm 和 scale 层，该层进行对输出特征的归一化。公式如下：

输入：x在mini$-batch$中的取值：$\mathrm{B}=\{\mathrm{x}_1...\mathrm{m}\}$;

学习的参数：τ，β

输出:$\{\mathrm{y}_\mathrm{i}=\mathrm{BN}_{\tau,\ \beta}(\mathrm{x}_\mathrm{i})\}$

$$\mu_\mathrm{s} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i$$

$$\sigma_\mathrm{S}^2 \leftarrow \frac{1}{\mathrm{m}}\sum_{i=1}^{m}(x_i-\mu_s)^2$$

$$\overline{\mathrm{x}_\mathrm{i}} \leftarrow \frac{x_i-\mu_s}{\sqrt{\sigma_\mathrm{S}^2+\varepsilon}}$$

$$\mathrm{y}_\mathrm{i} \leftarrow \tau\widehat{z}_i+\beta \equiv BN_{\tau,\ \beta}(z_i)$$

卷积层中，卷积核为 7*7*64，padding 深度为 3，stride 为 2，无 bias。

最大池化层中，池化核为 3*3，padding 深度为 0，stride 为 2。

resnet_2 层中，包括 a，b，c 层，每个分支又有两个 branch、卷积层、Eltwise 和 ReLU。

resnet_3 层中，包括 a，b，c，d 层，每个分支又有两个 branch、卷积层、Eltwise 和 ReLU。

resnet_4 层中，包括 a，b，c，d，e，f 层，每个分支又有两个 branch、卷积层、Eltwise 和 ReLU。

resnet_5 层中，包括 a，b，c 层，每个分支又有两个 branch、卷积层、Eltwise 和 ReLU。

均值池化层总，池化核为 7*7，padding 深度为 0，stride 为 1。

全连接层输出为 1*1*2。

3.2.3 训练 Resnet50 网络

使用实验室私有数据集和网络公开数据集（replay-attack、print-attack、ASCIA 等）在 caffe 平台上进行网络训练，私有数据集中包括 3200 真实人脸照片、904 彩色打印人脸照片、1141 手机屏幕人脸照片、410 电脑屏幕人脸照片，照片尺寸均为为 2448*3264，照片格式均为 jpg。

本课题使用随机梯度下降法（SGD）进行网络训练。

随机梯度下降法中，对于每次参数更新，仅使用一个数据来变换参数，这样虽然比完全梯度下降法的精度低，但是整体趋势仍是走向最小，而且算法较快，可以节省时间[16]。

随机梯度下降法的风险函数如下：

$$J(\theta)=\frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}(y^i-h_\theta(x^i))^2=\frac{1}{m}\sum_{i=1}^{m}\cos t(\theta,(x^i,y^i))$$

$$\cos t(\theta,(x^i,y^i))=\frac{1}{2}(y^i-h_\theta(x^i))^2$$

每个样本的损失函数如下：$\theta_j^{'}=\theta_j+(y^i-h_\theta(x^i))x_j^i$

3.3 结果验证和分析

本课题在 caffe 平台上训练 Resnet50 网络，使用 SGD 算法进行网络训练，迭代次数为 100000，使用实验室私有数据集进行训练，使用公开数据及 ASCIA、NUAA、print-attack 进行微调。结合 matlab 编写代码，输入为 224*224 像素的图片，识别结果为无摩尔纹则输出 0，识别结果为有摩尔纹则输出结果为 1。

在 replay-attack 数据集上进行测试时，准确率达到 99%以上。在实验室私有测试集中，包括 3200 真实人脸照片、904 彩色打印人脸照片、1141 手机屏幕人脸照片、410 电脑屏幕人脸照片，照片尺寸均为为 2448*3264，照片格式均为 jpg，在真实人脸照片集上准确率达到 100%、打印人脸照片集上准确率达到 100%、手机屏幕人脸照片集上准确率达到 100%、电脑人脸照片集上准确率从达到 100%。

不过在使用笔记本自带摄像头进行拍摄识别的时候准确率会有所下降，准确率在 95%左右，识别准确率会受到光照情况和是否带眼镜的影响。

4. 眨眼检测的研究

4.1 问题描述

如果计算机能够直接分辨出在摄像头前的是真实的人还是静止不动的照片、播放视频的电子设备，那么人脸识别系统必将更安全更可靠。而真实的人与静止不动的照片、播放视频的电子设备直接的最大的差距就是，人可以做各种动作，比如眨眼、皱眉，张嘴，也可以进行简单的人机交互，即跟随计算机指令做出相应动作。而这些照片、播放视频的电子设备却做不到。所以区分是真实的人还是静止不动的照片，可以通过判断摄像头前的用户，有没有做出一些只有真人能够做到而照片、电子设备做不到的行为。

眨眼检测是近几年发展起来的一种较为成熟的活体检测技术。通过判断用户有没有眨眼，再结合摩尔纹黑框检测，即可有效判断用户是否为真实的人。

4.2 研究方法

4.2.1 人脸特征点提取

本课题使用 SDM（Supervised Descent Method)）for face alignment 算法进行人脸特征点提取。

SDM 方法属于基于回归的方法[17]，基本原理是，处理一张待提取特征点的人脸时，首先给出一个初始的特征点的形状，再通过迭代运算，获取特征到偏移量的映射矩阵，将初始形状回归到接近于真实形状的位置。

训练 SDM 的映射矩阵步骤如下：

归一化样本，将样本归一化至 400*400 的大小。

计算均值人脸。

最小化函数：$\arg\min_{R_k,b_k}\sum_{d^i}\sum_{x_k^i}\left\|\Delta x_*^{k|i}-R_k\phi_k^i-b_k\right\|^2$，其中 $\phi_k^i=h(d^i(x_k^i))$，$\Delta x_*^{ki}=x_*^i-x_k^i$

使用实验室私有的手工标记的 4000 张图片进行训练获取特征到偏移量的映射矩阵，对齐效果如下图所示：

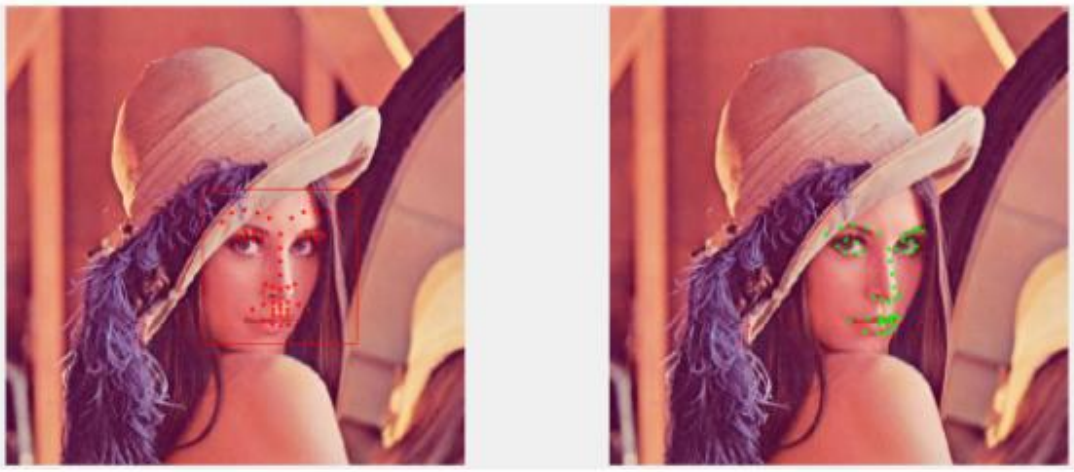

图 4.1 （a）49 个均值人脸特征点图（b）49 个人脸主要特征点图

本课题的算法输入为一张 RGB 图片，输出为标记了 49 个特征点图和 49 个特征点的坐标。

图中特征点的(x，y)坐标存储在一个 matlab 的 49*2 的 double 型数组内，其中 49 个特征点的存储顺序如下图：

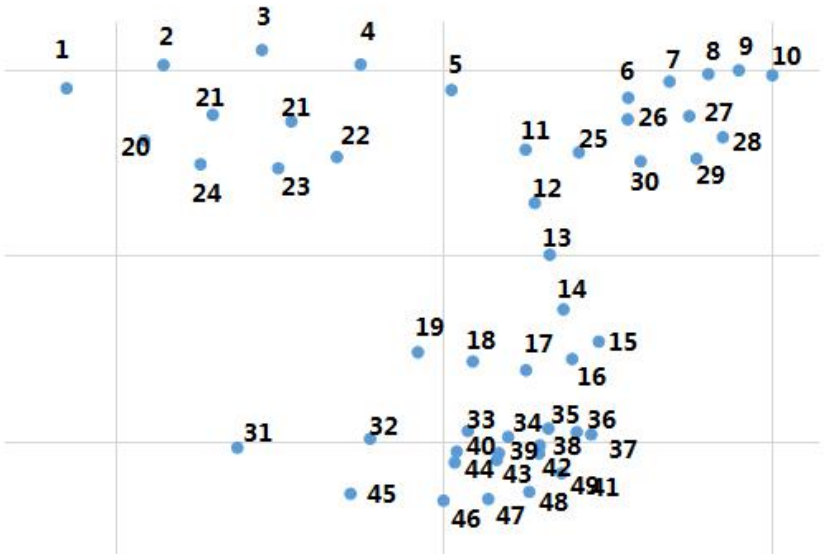

图 4.2 特征点存储顺序图

4.2.2 人眼结构模型

Tsuyoshi Moriyama[18]在 2002 年的论文中提出了对人眼进行建模，可以用来识别眼睛开合状态和检测眨眼次数。

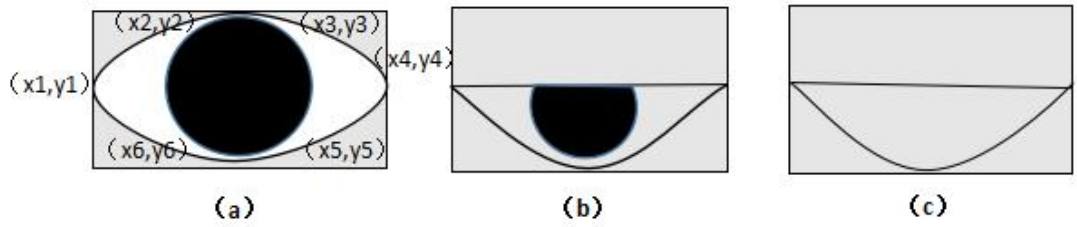

图 4.3 （a）睁眼图（b）半睁眼图（c）闭眼图

如图所示，定义眼睛开合度

$$U=\frac{\text{上眼皮最高点和下眼皮最低点距离}}{\text{内眼角和外眼角的距离}}\approx\frac{\frac{1}{2}(|(x2,y2)-(x6,y6)|+|(x3,y3)-(x5,y5)|)}{|(x1,y1)-(x4,y4)|}$$

可以看到，眼睛睁开时，眼睛开合度达到最大，闭眼时眼睛开合度达到最小，半睁眼时眼睛开合度介于睁开和闭上之间。

当眼睛睁到最大时，眼睛开合度 U 一般在 0.45 以上，闭眼时眼睛开合度为 0，定义眼睛闭合度 U 在 0.25 及以上为睁眼，U 在 0 到 0.25 之间为闭眼。

人眨眼会受到很多外界因素影响，当人处于安静状态时，眨眼的平均速度是 15 到 30 次每分钟，即每 2-4 秒眨一次眼，而完整的眨眼动作持续时间在 250 毫

秒左右。本课题使用笔记本自带摄像头进行拍摄，拍摄速度为 15 帧每秒，帧间间隔为 66 毫秒，即一次眨眼动作可以拍到 3-4 张图像。

所以认为只需要画出眼睛开合度时域变化曲线，同时统计眼睛开合度低于 0.25 的帧数就可以判断出眨眼次数。

4.2.3 眨眼检测

本课题眨眼检测是通过录一段每秒 15 帧的 5 秒视频进行检测，要求正视摄像头并且保证至少有两次眨眼。

使用的摄像头是笔记本自带摄像头，视频录制格式为 YUY2_640×480，视频保存格式为 avi 格式。

视频录制完毕后通过 Matlab 函数 read(obj,i)每两帧读取一次视频帧，提取视频帧的人脸的 49 个主要特征点坐标。其中，左上眼皮两个特征点为第 21,22 号点，右上眼皮两个特征点为第 27,28 号点，左下眼皮两个特征点为第 24,25 号点，右下眼皮两个特征点为 30,31 号点，左眼内眼角外眼角分别是 20,23 号点，右眼内外眼角点分别是 26,29 号点。

通过公式 $\frac{\frac{1}{4}(|21-25|+|22-24|+|27-31|+|28-30|)}{\frac{1}{2}(|20-23|+|26-29|)}$ 计算眼睛开合度，画出纵坐标为眼睛开合度，横坐标为视频帧的曲线，只要该曲线有两次低于 0.25，即可判断通过眨眼检测。

4.3 结果验证和分析

代码使用 matlab 编写完成，完成 15 帧每秒，共 5 秒 75 帧的视频识别大约需要 90 秒。

本课题分别录制了正脸戴眼镜眨眼的 10 个 5 秒视频、正脸不戴眼镜眨眼的 10 个 5 秒视频、侧脸戴眼镜眨眼的 10 个 5 秒视频、侧脸戴眼镜眨眼的 10 个 5 秒视频、正脸戴眼镜不眨眼的 10 个 5 秒视频、正脸不戴眼镜不眨眼的 10 个 5 秒视频、侧脸戴眼镜不眨眼的 10 个 5 秒视频、侧脸不戴眼镜不眨眼的 10 个 5 秒视频，一共 80 个 5 秒视频，每个视频中至少眨眼三次，使用本课题算法进行识别结果如下：

表 4.1 不同视频的识别结果表

视频	检测到不同眨眼次数的视频数				通过眨眼检测视频数
	一次及以上	两次及以上	三次及以上	四次及以上	
正脸戴眼镜眨眼	10	10	10	4	10
正脸不戴眼镜眨眼	10	10	10	5	10
侧脸戴眼镜眨眼	10	8	7	2	7
侧脸不戴眼镜眨眼	10	9	8	3	8
正脸戴眼镜不眨眼	0	0	0	0	0
正脸不戴眼镜不眨眼	0	0	0	0	0
侧脸戴眼镜不眨眼	0	0	0	0	0
侧脸不戴眼镜不眨眼	0	0	0	0	0

可以得出结论，在正视摄像头时，本课题的眨眼检测算法具有良好的识别率，而检测失败的时候大部分是因为待识别的用户侧脸面对摄像头，导致人脸特征点定位及特征点坐标输出产生一定误差。

5.结论与展望

5.1 结论

本课题使用 canny 算法、hough 变换完成了黑框检测 matlab 代码的编写，并且在实验室私有数据集上进行测试，识别准确率达到 97.2%。

本课题使用 caffe 平台，实验室私有数据集、网上公开数据集（ASCIA、NUAA、print-attack）使用随机梯度下降算法训练 Resnet50 网络，使用 Resnet50 网络完成摩尔纹检测 matlab 代码的编写，并且在实验室私有数据集、网上公开数据集 replay-attack 上进行测试，实验室私有测试集中，在真实人脸照片集上准确率达到 100%、打印人脸照片集上准确率达到 100%、手机屏幕人脸照片集上准确率达到 100%、电脑人脸照片集上准确率从达到 100%；replay-attack 数据集测试中，准确率达到 99%以上。

本课题使用 SDM for face alignment 算法，在 matlab 平台上训练人脸主要特征到偏移量的映射矩阵，编写眨眼检测代码，并且在个人视频数据集上进行测试，

在正视摄像头、光线良好的情况下准确率在 95%以上。

本课题使用 matlab 的 GUI 界面，将三种识别方法根据如下图流程图进行结合，先通过眨眼检测，再通过黑框检测和摩尔检测，将三种检测方法结合，形成本课题的活体检测算法，本课题编写的算法流程图如下：

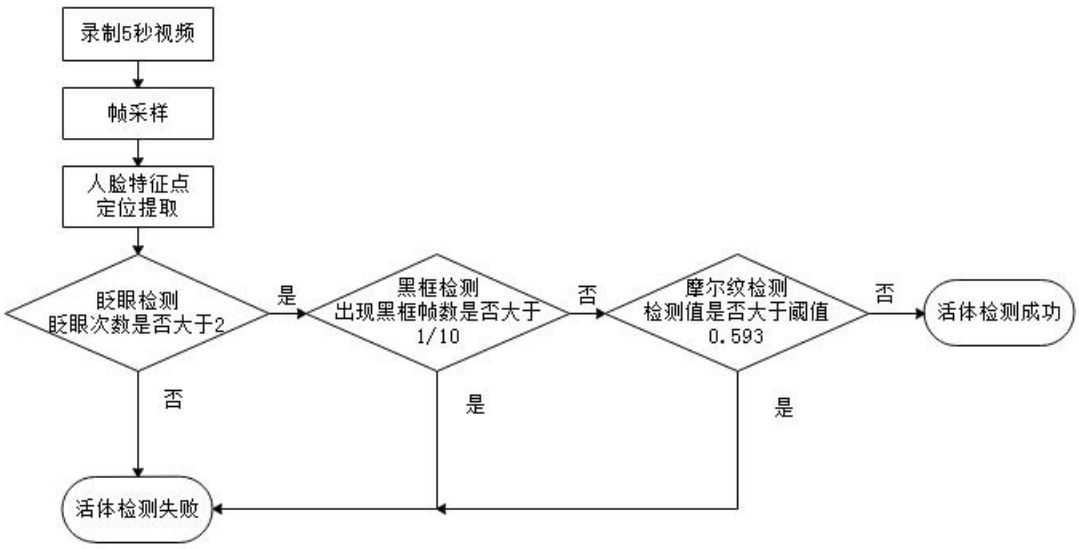

图 5.1 活体算法流程图

从图中可以看出，检测过程十分简洁，本算法从录制每秒 15 帧的 5 秒至少眨眼三次的视频开始，通过每两帧采样一帧图片的帧采样获取视频图片信息。再对帧采样获得的图片进行人脸特征点定位及特征点坐标提取。之后再进行眨眼检测，若对采样的图片进行眨眼检测得出的结果中眨眼次数小于等于两次则表示眨眼检测失败，活体检测结果为失败。

若大于两次则通过眨眼检测，进行第二步黑框检测，若采样的图片中检测出有大于等于采样数量 1/10 的图片中存在黑框，则表示黑框检测失败，活体检测结果为失败。

若黑框检测的存在黑框的图片数量小于采样数量 1/10，则通过黑框检测，进行第三步摩尔纹检测，对帧采样取得的每一张图片进行摩尔纹检测，若是每一张图片的摩尔纹检测结果均小于阈值 0.593，则判断为通过摩尔纹检测，通过活体检测，否则则判断活体检测失败。

本课题录制了十个视频对整体活体算法进行测试，结果如下：

A.眨眼三次，无黑框，无摩尔纹，5 秒活体真人视频；

通过眨眼检测，通过黑框检测，通过摩尔纹检测，通过活体检测；

B.不眨眼，无黑框，无摩尔纹，5 秒活体真人视频；

不通过眨眼检测，活体检测失败；

C.眨眼三次，无黑框，无摩尔纹，5 秒手机屏幕真人视频；

通过眨眼检测，通过黑框检测，通过摩尔纹检测，通过活体检测；

D.不眨眼，无黑框，无摩尔纹，5 秒手机屏幕真人视频；

不通过眨眼检测，活体检测失败；

E.眨眼三次，有黑框，无摩尔纹，5 秒手机屏幕真人视频；

通过眨眼检测，不通过黑框检测，活体检测失败；

F.不眨眼，有黑框，无摩尔纹，5 秒手机屏幕真人视频；

不通过眨眼检测，活体检测失败；

G.眨眼三次，无黑框，有摩尔纹，5 秒手机屏幕真人视频；

通过眨眼检测，通过黑框检测，不通过摩尔纹检测，活体检测失败；

H.不眨眼，无黑框，有摩尔纹，5 秒手机屏幕真人视频；

不通过眨眼检测，活体检测失败；

I. 眨眼三次，有黑框，有摩尔纹，5 秒手机屏幕真人视频；

通过眨眼检测，不通过黑框检测，活体检测失败；

J.不眨眼，有黑框，有摩尔纹，5 秒手机屏幕真人视频；

不通过眨眼检测，活体检测失败；

可以得出结论，算法测试结果良好，只在测试眨眼三次，无黑框，无摩尔纹，5 秒手机屏幕真人视频的时候出现了失误，算法对于高质量、无黑框、无摩尔纹的视频欺骗抵抗能力较弱，不过这种欺骗比较难以做到，所以在接受范围内。

5.2 论文中出现的问题及思考

本课题的目标是做出一个用户体验良好的活体检测算法，不需要外加设备，不需要过多的人机交互，能够有效抵抗外界干扰，所以从众多活体检测线索中挑选可以作为本课题的线索的时候出现了很多问题。

一开始有考虑人机交互较低的人脸三维深度信息作为活体检测线索之一，但是单个普通摄像头无法获得人脸三维深度信息，需要双目摄像头或者是体感摄像头 kinect，这有悖于本课题的初衷，所以放弃了这个线索。

也有考虑过使用嘴部特征点来代替眨眼检测，但是嘴部特征的这个线索需要较多的人机交互，或是让用户说一句特定的话或是张嘴微笑或是什么，也有悖于减少人机交互的初衷，所以放弃，而眨眼检测所需要的人机交互远远低于嘴部的特征点，具有良好的用户体验。

最初看到摩尔纹检测这条活体检测线索是在了解 linkface 的活体检测方案的时候，该方案提到现在做摩尔纹检测的人还不多，而且摩尔纹检测不需要人机交互，可以在用户不知情的情况下进行，可以降低非法用户通过了解活体检测而针对欺骗的可能性。

选择黑框检测是因为，目前的人脸活体检测欺骗手段主要是打印照片、手机屏幕照片、电脑屏幕照片、手机屏幕视频等，而我们认为在使用这些手段对着摄像头进行活体检测的时候，画面中或多或少会出现黑框，而且黑框检测算法简介，运行速度快，可以高效率得降低被欺骗的概率。

不过在本课题完成的过程中，同样发现了许多问题，第一个是在黑框检测时，会受到外界干扰如光照的影响，光照情况不太良好时，手机屏幕边框和照片的边界差异没有那么明显，会降低黑框检测的识别率；第二个是摩尔纹检测时，并不是所有的手机屏幕照片、电脑屏幕照片都会产生摩尔纹，具有一定的偶然性；第三个是眨眼检测由于是需要录制视频之后进行视频帧的检测，在程序运行的时候大部分时间花在了帧检测上面，而且等待时间比较长，拖慢了整体活体检测的时间，检测时间方面还有待提高。

5.3 展望

本课题进行活体检测主要用到了三个线索：黑框检测、眨眼检测、摩尔纹检测。这三个线索对于抵抗照片、视频重放攻击、3D 人脸模型攻击、外界干扰的能力并不是最优选择。这三个线索对于照片攻击的抵抗能力较强，对于视频攻击的抵抗能力一般，对于 3D 人脸模型攻击的能力较弱，对于外界干扰的抵抗能力一般。且结合起来代码运行，检测速度较慢，还有待于提高。

目前活体检测仍处于刚起步的阶段，欺骗手段相对比较简单单一，所以本课题的算法目前来说仍是良好的算法。不过目前网络上已经出现了使用一张照片，制作出同一个人不同表情的软件，相信未来不久这种软件，会导致活体检测的欺骗手段多样化，会变得更加难以抵抗。

在挑选活体检测线索的时候，本课题了解到三维深度信息、人体红外信息是最能够反应待检测用户是否为活体的信息，而且是对各种欺骗手段抵抗能力最好的两种线索，所以之后可以将研究热点定在三维深度信息和人体红外信息上。

不过，三维深度信息的获取难度，红外信息获取的设备价格都是限制这两个线索使用的重要因素。所以本课题觉得如何通过单张照片提取照片的深度信息、如何降低红外信息获取设备的价格都是很好的研究方向。

相信未来的人脸识别活体检测技术，能够达到一个非常高的高度，能够实时的、快速的、对于欺骗抵抗能力相当高的、对于外界干扰抵抗能力强的、人机交互少、用户体验良好的活体检测技术，让人脸检测技术变得更加安全，让人脸成为每个人独一无二的信息，可以通过人脸信息来完成各种领域的交互。

参考文献

[1]王馨宁.生物特征识别中的“活体检测”概念及分析.[A].国家知识产权局专利局专利审查协作中心.2014-6-14

[2]方植彬.信息与通信网络安全技术——生物识别技术[J].电子产品可靠性与环境试验,2014,32(05):55-61.

[3] 天诚盛业. 回顾 20 年生物识别技术发展艰辛路.http://blog.sina.com.cn/s/blog_8693e9330102wvc2.html.2017-1-19

[4] Choudhury.T，Clarkson.B，Jebara.T etc. Multimodal person recognition using unconstrained audio and video.AVBPA’99，1999，176~181

[5]Thalheim.L，Krissler.J，Ziegler.PM. Body Check: Biometric Access Protection Devices and their Programs Put to the Tese. c’t magazine，November 2002

[6] 孙霖. 人脸识别中的活体检测技术研究[D].浙江大学,2010

[7] Choudhury.T，Clarkson.B，Jebara.T etc. Multimodal person recognition using unconstrained audio and video.AVBPA’99，1999，176~181
[8] Kollreider.K，Fronthaler.H，Bigun.J. Evaluting liveness by face images and the structure tensor. Fourth IEEE Workshop on Automatic Identification Advanced Technologies，2005，75~80
[9] 许晓. 基于深度学习的活体人脸检测算法研究[D].北京工业大学,2016
[10] Chetty.G，Wagner.M.，Liveness Verification in Audio-Video Speaker Authentication. 8th International conference on Spoken Language Processing，2004
[11] Li Jiangwei，Wang Yunhong，Jain A K. Live Face Detection Based on the Analysis of Fourier Spectra. Biometric Technology for Human Identification，Proc. Of SPIE，2004,404:296~303
[12]李翼. 应用于人脸识别中的反照片欺骗检测方法研究[D].南京航空航天大学,2011
[13]知乎话题-摩尔纹.https://www.zhihu.com/topic/19822099/hot
[14] 蒋竺波. 知乎回答《如何理解微软的深度残差学习》.https://www.zhihu.com/question/38499534
[15]Kaiming He，Xiangyu Zhang，Shaoqing Ren etc. Deep Residual Learning for Image Recognition.CVPR,2016 IEEE Conference on
[16]金钊.基于改进随机梯度下降算法的SVM[D].河北大学,2017.
[17] 贾金让. 人脸对齐之SDM论文解析.https://blog.csdn.net/wfei101/article/details/73257851
[18]]Moriyam.T，Kanade.T，Cohn.J.F. etc. Automatic Recognition of Eye Blinking in Spontaneously Occurring Behavior.Proceedings of the 16th International Conference on pattern Recognition，2002(4):78-81