**使用闪光灯的反2D攻击面部活性检测**

**论文题目：Face Liveness Detection Using a Flash Against 2D Spooﬁng Attack**

## 介绍及研究意义：

当今社会人脸识别已经得到了广泛应用，然而传统的人脸识别系统往往没有考虑恶意攻击者的存在。许多研究都表明，这样的系统极易受到欺骗攻击，即攻击者伪装成授权的人，获得对系统的非法访问。

比较典型的例子是**二维欺骗攻击**，它通过使用有效用户的二维面部副本来迷惑系统，是一种最常见的攻击方法，因为一个人的图像或视频是很容易获取和高度可复制的。二维欺骗攻击有三种类型，即**照片攻击**、**视频攻击**和**模拟面具攻击**。照片攻击通过在一张纸上或电子屏幕上使用合法用户的照片来逃避检测；而视频攻击通过在电子设备上通过播放授权人员的视频来迷惑系统；在模拟面具攻击中，对手戴着一个2D面具伪装成授权人员。

人脸活跃度检测，又称人脸欺骗检测，是为了抵御二维欺骗攻击而设计的检测手段。人脸活度检测可以在人脸识别过程开始前确定图像是来自真实的还是虚假的主体。可疑图像会被过滤，而不会被传送到识别系统。

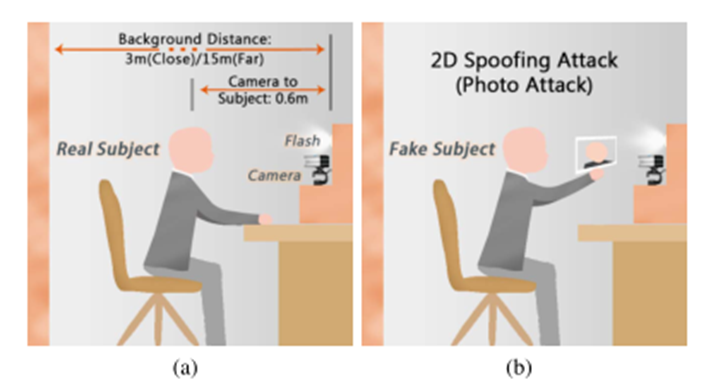
## 研究现状：

人脸活度检测可分为**基于硬件**和**基于软件**两种方式。以往人脸活度检测的工作主要集中在基于软件的方法上，通过分析被试的纹理、结构信息、活度标志以及所捕获图像的质量等数据来进行判别检测。这样的方法通常对环境因素十分敏感，如光照条件、图像噪声等。因此在这种情况下，其检测精度会明显下降。此外，某些活动性线索的计算复杂度较高，如基于连续帧的人脸动态计算。虽然要求用户说话或摇头可以提高检测的准确性，但由于检测时间较长且用户不一定配合，也会降低检测效率。

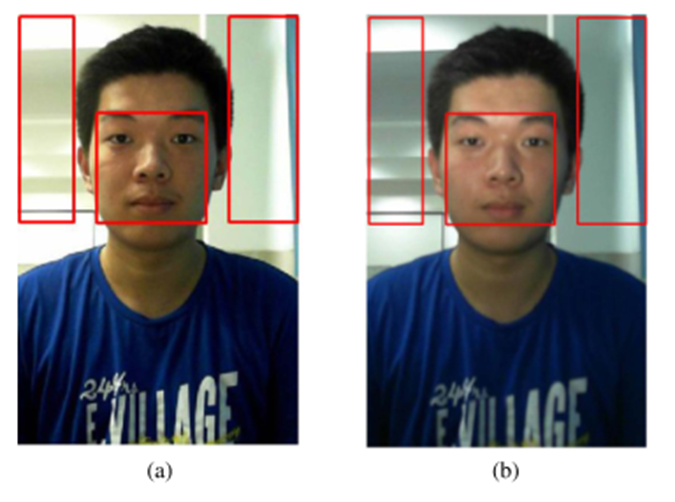
还有一种工作是在基于硬件的识别系统中嵌入一个设备来获取受试者的额外信息，例如体表温度等。尽管如此，一些额外的硬件是较为昂贵而且复杂难以安装的。

## 本文方法：

本文阐述一个利用flash进行人脸活性检测的模型，以抵御照片、视频以及模拟掩膜攻击。



在该模型中，检测过程中从检测对象上获取一对图像，一幅没有闪光（如下图(a)），另一幅使用闪光（如下图(b)）。利用基于**均匀局部二值模式**的描述符对人脸纹理信息进行测量，然后利用同一检测对象带flash图像和不带flash图像的**标准差**和**灰度差均值**获取人脸结构信息的三种描述符。该方法的特征值是经过精心设计的，以提供更准确、鲁棒性更强、并有低时间复杂度的预测。

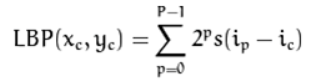


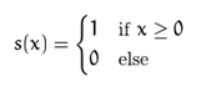
首先确定人脸区域 **𝐼\_𝑛^𝐹** ；然后是基于人脸区域定位两个背景区域，记为 **𝐼\_𝑛^𝐵𝐺**

下面主要介绍一下四个描述符：

1. **Flash图像描述符(LBP\_FI)**

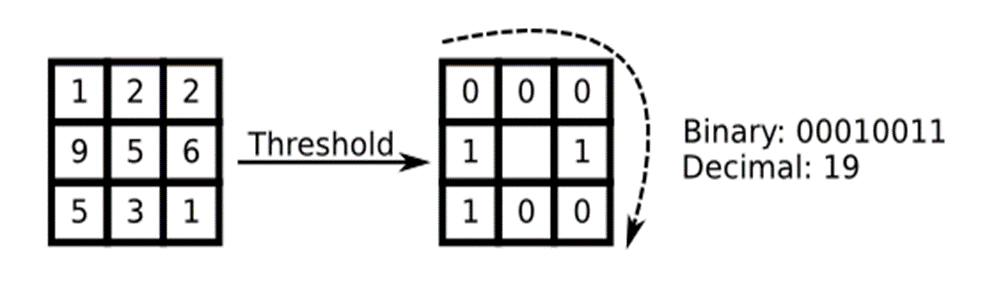
首先将人脸区域  分成9个不重叠的块，获取图像不同区域的纹理信息。然后计算每个块中像素的LBP码。





表示中心的像素点，为其灰度值。

以下图为例：



中心点灰度值为5，该点的LBP值为19。

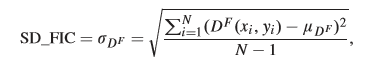
由于采用传统LBP模式的统计直方图来表达图像的信息时，较多的模式种类将使得数据量过大（P个采样点的LBP将会产生种模式），且直方图过于稀疏。因此，需要对原始的LBP模式进行降维，使得数据量减少的情况下能最好的代表图像的信息。为了解决二进制模式过多的问题，提高统计性，Ojala提出了采用一种“等价模式”来对LBP算子的模式种类进行降维，所以在我们的模型中，使用的是降维的**等价LBP（Uniform LBP）**。

当某个LBP所对应的循环二进制数从0到1或从1到0最多有两次跳变时，该LBP所对应的二进制就称为一个等价模式类。如00000000（0次跳变），00000111（只含一次从0到1的跳变），10001111（先由1跳到0，再由0跳到1，共两次跳变）都是等价模式类。除等价模式类以外的模式都归为另一类，称为混合模式类，例如10010111（共四次跳变）。

通过这样的改进，二进制模式的种类大大减少，而不会丢失任何信息。模式数量由原来的减少为 P ( P-1)+2。对于3×3邻域内8个采样点来说，二进制模式由原始的256种减少为58种，这使得特征向量的维数更少，并且可以减少高频噪声带来的影响。

这个描述符的作用是用来从图片的纹理等信息上来进行判断，即类似于传统的基于软件的检测方法。

1. **人脸亮度变化的标准差描述符(SD\_FIC)**



SD\_FIC测量由flash引起的人脸区域灰度强度变化。由于闪光的结构信息不同，其在真实人脸中的反射也不同，即闪光与人脸各部分之间的距离可能不同（二维欺骗攻击往往是一个平面图像）。相比之下，二维欺骗攻击的反射光更加均匀。因此，真人攻击强度的偏差大于二维欺骗攻击强度的偏差。利用标准差来捕捉模型中灰度强度的变化。

和 分别表示  的均值和标准差，后面也是类似的表达。N是该区域的像素数量。由于三维物体上的强度变化明显区别与二维，真实人脸的SD\_FIC值在所有情况中是最大的。纸照片、二维掩码攻击和曲面掩码攻击由于会在人脸区域出现明亮的条带，因此比其他类型的攻击具有更大的SD\_FIC。

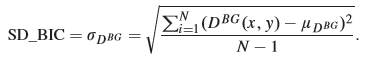
1. **背景强度变化平均值描述符(M\_BIC)**



这个描述符主要针对的是背景区域。实际背景在图片攻击和视频攻击中会被阻挡。由于从显示媒体上捕获的背景往往比实际的背景离相机更近，因此会反射出更高强度的光。借此设计M\_BIC来捕捉这些信息。

由于真实人脸和两种遮罩攻击都不会遮盖真实背景，真实背景在使用了flash的图像中就会被flash遮挡，这样一来，照片和视频攻击的M\_BIC值要大得多，真实的面孔和两种面具攻击就会具有相近的M\_BIC值。

1. **背景强度变化标准差描述符(SD\_BIC)**



与之前的照片和视频攻击不同，在二维掩码攻击或曲面掩码攻击中，只使用受试者头部区域，不涉及实际背景。由于纹理和形状不同，其面具的光扩散效果也不同于真实的脸。

SD\_BIC描述符主要利用针对背景区域的对比。真实人脸的SD\_BIC值会小于掩码攻击的SD\_BIC值。这是因为面罩的光扩散要比人脸的光扩散大。由于iPad和照片的二维平面结构，flash在iPad和纸质照片攻击中均匀增加了背景区域的强度。这些攻击的SD\_BIC会比不覆盖真实背景的攻击要小。

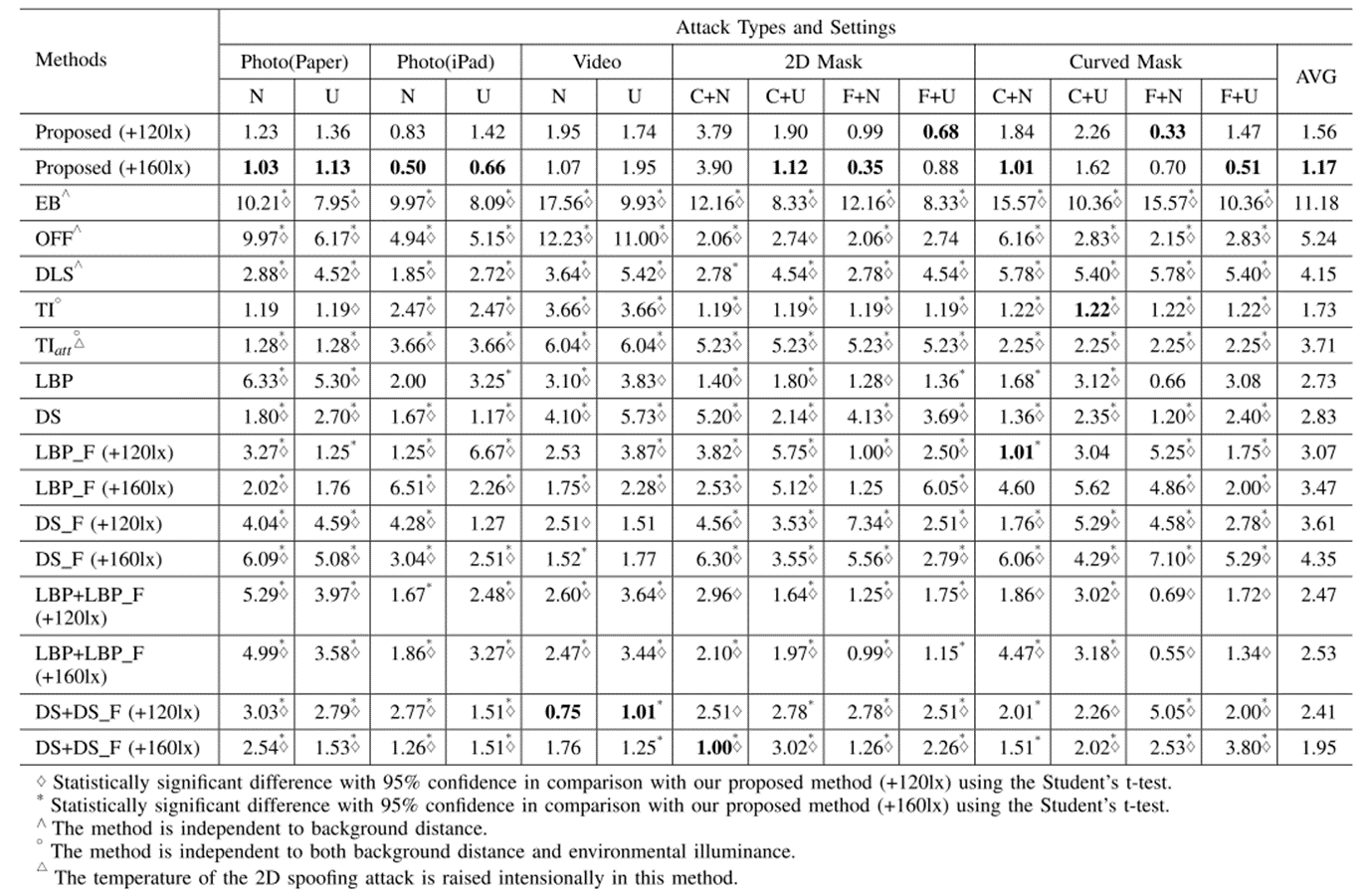
**真实人脸和不同攻击对应的结构描述符**

|  |  |  |  |
| --- | --- | --- | --- |
|  | SD\_FIC | M\_BIC | SD\_BIC |
| 真实人脸 | 39.45 | 36.88 | 24.02 |
| 纸照片 | 19.42 | 62.12 | 25.81 |
| 电子照片 | 18.52 | 58.87 | 17.13 |
| 视频 | 17.03 | 63.24 | 13.11 |
| 2D面具攻击 | 30.44 | 35.57 | 37.76 |
| 卷曲面具攻击 | 33.80 | 43.88 | 33.88 |

在本文提出的模型中，LBP\_FI描述符捕获纹理信息，而SD\_FIC、M\_BIC和SD\_BIC测量结构信息，四个描述符都有针对的攻击领域，综合起来可以区别各种二维攻击。这就解释了为什么我们的方法可能比没有flash的方法更准确。此外，由于flash与环境光相比具有较强的照度，减少了环境光照程度的影响，因此对其活度的检测更加稳定。

## 实验结果及分析：

本文通过运行时间和常用的半总错误率(HTER)准则来评价各种活度检测方法的性能的比较，HTER越小说明表现越好。论文中给出了本方法和其他现有检测方法的HTER，加粗的数字为列最小：

****

从上图可以看出，基于flash的检测技术在照片攻击和模拟面具攻击上表现为最优，在视频攻击上也极为接近最优水平。而且其他方法大多收到各种条件制约。

## 总结展望：

本文提出了一种基于flash的人脸活性检测二维欺骗攻击的方法。根据四种描述符测量的带flash和不带flash图像的差异，将被试分为合法类和恶意类。与基于硬件的方法不同，我们的方法只需要闪光灯，既经济又易于安装在现有的人脸识别系统中。由于flash增强了真假人脸的差别，减少了环境光照的影响，因此该方法比基于软件的方法更准确、鲁棒性更强。此外，提取四个描述符的时间复杂度较低，不需要用户协作。此方法既利用了基于软件的方法，又利用了基于硬件的方法。具有精度高、鲁棒性强、计算复杂度低、安装成本低等优点。

本文的方法也有一定的局限性，比如对闪光灯的距离有严格要求，而且闪光灯的位置要对用户友好，虽然我们目前型号的闪光灯的照度对人眼没有伤害，也远低于相机使用闪光灯的照度，但是用户的舒适度是一个问题。克服这一限制的一种可能的解决办法是调整闪光角度。如果不将闪光灯安装在眼睛的高度，闪光灯的照射就不会直接刺激人眼，反而会让人感觉更舒适。闪光角度的确定不仅要根据检测精度，还要根据安装难度。

这一研究未来的工作可能是专注于探索模型的性能，使之可以对更先进的攻击进行检测防御，比如3D欺骗攻击，即精良的3D人脸面具和3D模型的各种表情。由于表面反射率的不同，真实人脸和3D掩膜的反射光是不同的，此外3D的纹理细节也可以通过flash来增强。因此，如果能够找出合适的描述符，额外的flash照明也应该有助于将合法用户与恶意攻击区分开来。