

DOI:10.16356/j.1005-2615.2020.05.012

## 基于光照方向一致性的换脸视频检测

李纪成<sup>1</sup>, 刘琲贝<sup>1,2</sup>, 胡永健<sup>1,2</sup>, 王宇飞<sup>2</sup>, 廖广军<sup>3</sup>, 刘光尧<sup>4</sup>

(1. 华南理工大学电子与信息学院, 广州, 510641; 2. 中新国际联合研究院, 广州, 510000;  
3. 广东警官学院刑事技术系, 广州, 510230; 4. 公安部物证鉴定中心, 北京, 100038)

**摘要:** 光照条件是视频外部成像环境的重要因素, 可反映视频成像时的客观物理条件, 即使采用复杂的计算机图形学模型, 也难以完全反映光照的真实场景。针对大部分 Deepfake 换脸视频与真实视频在外部成像环境上的差异, 提出一种基于光照方向一致性的换脸视频检测方法。本文利用 Lambert 光照模型逐帧计算待测视频的二维光照方向, 通过判断整段待测视频二维光照方向的角度变化是否平滑确定视频真伪。实验结果表明该方法在公开的测试数据库 TIMIT 和 FaceForensics<sup>++</sup> 上具有较好的检测性能, 可以有效区分真实视频和换脸视频。由于不需要训练检测模型, 该方法具有计算复杂度低、实时性好的优点。

**关键词:** 视频篡改检测; 换脸视频; 成像环境; 光照方向

**中图分类号:** TP391.41      **文献标志码:** A      **文章编号:** 1005-2615(2020)05-0760-08

## Deepfake Video Detection Based on Consistency of Illumination Direction

LI Jicheng<sup>1</sup>, LIU Beibei<sup>1,2</sup>, HU Yongjian<sup>1,2</sup>, WANG Yufei<sup>2</sup>, LIAO Guangjun<sup>3</sup>, LIU Guangyao<sup>4</sup>

(1. School of Electronic and Information Engineering, South China University of Technology, Guangzhou, 510641, China;  
2. Sino-Singapore International Joint Research Institute, Guangzhou, 510000, China;  
3. Faculty of Forensic Science and Technology, Guangdong Police College, Guangzhou, 510230, China;  
4. Institute of Forensic Science, Ministry of Public Security, Beijing, 100038, China)

**Abstract:** Illumination is an important part of the external imaging environment of a video, which reveals the physical conditions under which the video is taken. It is difficult to simulate the realistic illumination even using complicated computer graphic models. Based on the difference of external imaging environment between real videos and Deepfake videos, we propose a method for detecting Deepfake videos by exploiting the consistency of illumination directions. Specifically, we employ the Lambert illumination model to calculate the 2-D illumination directions of the video on a frame-by-frame basis. The authenticity of the video is determined by examining the smoothness of direction change of the entire video. Experiments on the public test datasets TIMIT and FaceForensics<sup>++</sup> show that our method can effectively distinguish real videos from Deepfake videos. The method features low computational complexity since it does not involve any model training stages, making it more appropriate for real-time applications.

**Key words:** video forgery detection; deepfake video; imaging environment; illumination direction

随着人工智能和机器学习的快速发展, 视频换脸的技术门槛变得越低。从早期基于电脑端的 Deepfake 视频换脸技术<sup>[1-2]</sup>, 到目前基于手机端的 Impressions 换脸 APP<sup>[3]</sup>, 视频换脸技术对于硬件

**基金项目:** 科技部“前沿科技创新”专项(2019QY2200)资助项目; 中新国际联合研究院(206-A017023, 206-A018001)资助项目; 广州市产业技术重大攻关计划(201902010028)资助项目; 广州开发区科技计划(2017GH22)资助项目; 广东省自然科学基金博士科研启动(2017A030310320)资助项目; 中央高校基本科研业务费专项资金(2019MS025)资助项目; 广东省教育厅特色创新类(2017KTSCX132)资助项目。

**收稿日期:** 2020-07-15; **修订日期:** 2020-08-19

**通信作者:** 刘琲贝, 女, 讲师, E-mail: eebbliu@scut.edu.cn。

**引用格式:** 李纪成, 刘琲贝, 胡永健, 等. 基于光照方向一致性的换脸视频检测[J]. 南京航空航天大学学报, 2020, 52(5): 760-767. LI Jicheng, LIU Beibei, HU Yongjian, et al. Deepfake video detection based on consistency of illumination direction[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(5): 760-767.

的要求越来越低,且换脸视频的制作时长也从最初的几个小时变成了短短几分钟,使用者仅通过一部手机就可以在几分钟内得到效果逼真的换脸视频。然而这些换脸视频在快速传播的同时,也给社会舆论带来了极大的安全隐患,因此针对换脸视频篡改检测技术的研究具有重要的意义。

换脸视频丰富了人类的社会生活,也带来了巨大的安全隐患,引起了社会各界的广泛关注,目前国内外学者已对换脸视频篡改检测技术展开相关研究<sup>[4]</sup>。现今流行的换脸视频篡改检测技术主要通过深度神经网络判别器来检测视频是否发生换脸<sup>[5-14]</sup>。在库内的检测准确率通常都较高,但跨库性能严重下降,泛化性能或称通用能力不强,这也是目前基于深度神经网络检测算法的共性问题。此外,基于深度网络的检测方法需要预先进行模型训练,使得算法计算复杂度较高,无法在普通计算机上直接运行。除此之外,在实现的技术细节方面,脱胎于针对图像处理而设计的深度网络检测算法<sup>[5-13]</sup>多以帧图像为单位进行模型训练和检测,未能考虑同一视频相邻帧之间在时间轴上的内在联系,即时域特征,虽有个别算法<sup>[14]</sup>考虑了连续若干帧在时域上的特征关联,但计算复杂度高,特征描述针对性不强。

为了降低计算复杂度,提高通用性,本文从视频成像的基本事实出发,利用常规图像处理技术,提出一种不依赖于深度网络特征且简单可行的方法。由于真实视频和换脸视频在成像环境上存在差异,一种合理的途径是利用这种差异进行篡改检测。真实视频一般直接通过数码成像设备录制得到;而换脸视频则是将换脸处理过的视频帧序列重新进行视频编码得到,二者在成像方式上有本质不同。

光照方向是一个重要的取证特征<sup>[15-18]</sup>,能客观反映成像环境,故从光照特征出发来分析真实视频和换脸视频之间的差异具有天然的合理性。一般而言,用于换脸的目标视频通常为在自然环境中拍摄的短视频,这类真实视频相邻帧之间的光照方向变化具有较好的一致性;而换脸视频在对单帧图像进行换脸时虽考虑了光照渲染,但由于各换脸帧常为独立制作,难以保证相邻帧图像之间光照方向的自然吻合。

本文提出一种基于光照方向一致性的换脸视频篡改检测算法。首先确定待测视频的脸部检测区域,然后逐帧计算检测区域的二维光照方向矢量,以获得整段待测视频的光照角度序列,最后将

光照角度序列的标准差与预先设定的判别门限进行比较,确定待测视频是否换脸。在公开的换脸视频测试数据库 TIMIT 和 FaceForensics<sup>++</sup> 上的实验结果表明,本文算法可以快速有效地区分真实视频和换脸视频,与现有 3 种算法相比,也具有最好的综合检测性能。

## 1 光照方向的计算

Lambert 光照模型<sup>[19-20]</sup>(图 1)是一种流行的单光源漫反射模型,假设光源是单一无限远的点光源,且物体表面会等方向性地反射光线,则观察者在任何一个方向观察到的反射亮度都相同。

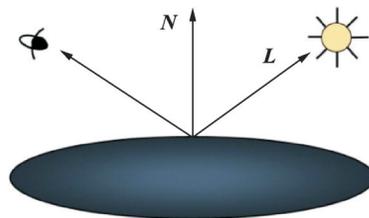


图1 Lambert光照模型

Fig.1 Lambert illumination model

虽然不是所有光照环境都是单一点光源,也不是所有物体表面都会等方向性地反射光线,但 Lambert 模型仍是一个简单通用的光照模型,本文沿用这个模型。物体表面反射光的光照强度  $I$  的定义如下

$$I = k(\mathbf{N} \cdot \mathbf{L}) + \epsilon \quad (1)$$

式中: $k$ 为物体表面的漫反射率,经验值为 $k=1$ ;  $\mathbf{N}$ 为物体表面的法线方向矢量; $\mathbf{L}$ 为照射到物体表面上的光照方向矢量; $\epsilon$ 为环境光强度。由于实际的光照方向是三维矢量,物体表面的理论光照强度值可通过式(1)得到,实际光照强度值可用物体表面的灰度像素值表示,则三维光照方向矢量 $\mathbf{L}$ 可通过最小化理论光照强度和实际光照强度的差异得到。其中理论光照强度和实际光照强度的差异通过二次误差函数  $E$  描述, $E$  的定义如式(2)和式(3)所示。

$$E(\mathbf{L}, \epsilon) = \left\| \mathbf{M} \begin{pmatrix} L_x \\ L_y \\ L_z \\ \epsilon \end{pmatrix} - \begin{pmatrix} I(x_1, y_1) \\ I(x_2, y_2) \\ \vdots \\ I(x_p, y_p) \end{pmatrix} \right\|^2 = \|\mathbf{M}\mathbf{v} - \mathbf{b}\|^2 \quad (2)$$

$$\mathbf{M} = \begin{pmatrix} N_x(x_1, y_1) & N_y(x_1, y_1) & N_z(x_1, y_1) & 1 \\ N_x(x_2, y_2) & N_y(x_2, y_2) & N_z(x_2, y_2) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ N_x(x_p, y_p) & N_y(x_p, y_p) & N_z(x_p, y_p) & 1 \end{pmatrix} \quad (3)$$

式中: $\|\cdot\|$ 表示向量范数, $L_x, L_y$ 和 $L_z$ 表示三维光照方向矢量 $L$ 在三维空间坐标系 $x, y, z$ 方向上的分量, $p$ 表示用于计算光照方向的采样点数量, $N_x(x_i, y_i), N_y(x_i, y_i)$ 和 $N_z(x_i, y_i)$ 表示第 $i$ 个采样点处的三维法向矢量在 $x, y, z$ 方向上的分量, $I(x_i, y_i)$ 表示在第 $i$ 个采样点处的实际光照强度值。通过对式(2)的二次误差函数求导,使其最小化,从而得到三维光照方向 $L_0, L_0$ 表达式为

$$L_0 = \begin{pmatrix} L_x \\ L_y \\ L_z \\ \epsilon \end{pmatrix} = (M^T M)^{-1} M^T b \quad (4)$$

实际应用场景中通常针对图像计算其二维光照方向,此时可令 $L_0$ 中 $z$ 分量等于0得到二维光照方向 $L_1, L_1$ 表达式为

$$L_1 = \begin{pmatrix} L_x \\ L_y \\ \epsilon \end{pmatrix} = (M_1^T M_1)^{-1} M_1^T b \quad (5)$$

$$M_1 = \begin{pmatrix} N_x(x_1, y_1) & N_y(x_1, y_1) & 1 \\ N_x(x_2, y_2) & N_y(x_2, y_2) & 1 \\ \vdots & \vdots & \vdots \\ N_x(x_p, y_p) & N_y(x_p, y_p) & 1 \end{pmatrix} \quad (6)$$

式中矩阵 $M_1$ 由矩阵 $M$ 去除其第3列元素后得到。

## 2 本文算法原理

### 2.1 算法原理

首先用一段帧长为81的真实视频和其对应的换脸视频来介绍本文方法的基本思想(图2)。



注:第一行为真实视频,第二行为换脸视频,从左到右依次为视频的第1,4,7,10和13帧

图2 真实视频和其对应的换脸视频帧图像

Fig.2 Frames of a real video and its corresponding fake video

图3为图2中真实视频和其对应换脸视频光照方向角度分布的对比图,由图3可以看出,前者在时间轴上变化较平缓,而后者在时间轴上变化较剧烈。对于一段数十秒的短视频,可以假设拍摄环境的光照条件是不变的,即在自然环境中拍摄的真实视频其相邻帧之间光照方向的变化具有较好的一致性;与之相反,换脸视频则是通过将独立换脸

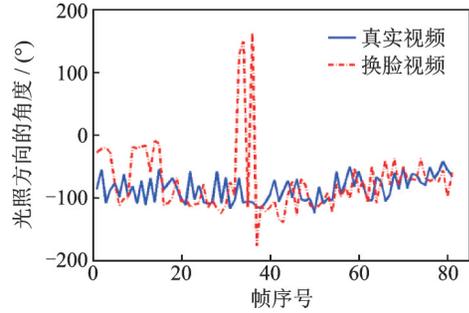


图3 真实视频和其对应换脸视频光照方向角度对比

Fig.3 Comparison of illumination direction distribution between a real video and its corresponding fake video

的视频图像帧序列重新编码得到,独立换脸过程难以保证相邻帧之间光照方向的自然吻合。据此,当待检测视频帧序列的光照方向角度出现剧烈变化时,有理由认为该段视频可疑,若光照方向的计算区域限定在脸部,则可认为是发生了换脸操作。

### 2.2 算法步骤

本文算法通过视频帧图像的纹理信息估计其二维光照方向,主要包括视频分帧、选择检测区域、选择目标区域和计算目标区域二维光照方向4个部分,其算法流程如图4所示。

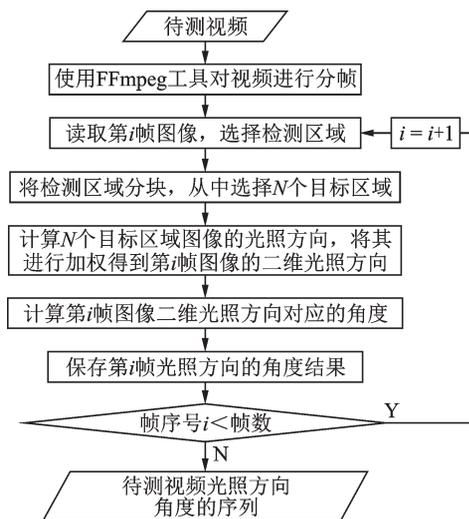


图4 计算视频光照方向序列的算法流程图

Fig.4 Flow chart for calculating illumination direction sequence of a video

下面以图2中的真实视频作为测试对象说明本文算法的计算过程。具体步骤如下:

**步骤1** 逐帧读取视频帧图像,并使用Dlib从视频帧图像中提取人脸部分作为检测区域。

**步骤2** 将各帧检测区域图像转化到YCbCr图像空间,提取其对应的灰度分量图像 $I_G$ ,然后利用Canny边缘检测算子对 $I_G$ 进行边缘滤波,得到其对应的二值边缘图像 $I_C$ 。

**步骤3** 将 $I_G$ 和 $I_C$ 等比例划分为 $4 \times 4$ 个子区

域,其中子区域序号  $k$  按照逐行扫描顺序排列,通过  $I_G$  计算每个子区域的平均像素值  $G_k$ ,通过  $I_C$  计算每个子区域的边缘信息所占百分比  $P_k$ ;以测试视频第 1 帧图像为例,其对应的  $G_k$  和  $P_k$  分布情况如图 5 所示。

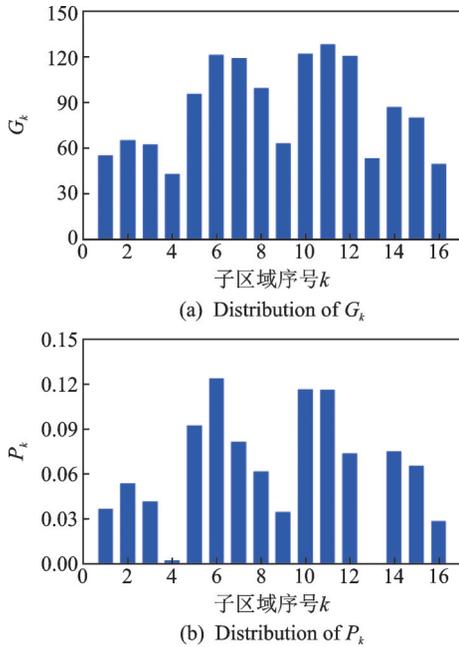


图 5 测试视频第 1 帧图像对应的  $G_k$  和  $P_k$  分布示意图  
Fig.5 Distribution of  $G_k$  and  $P_k$  corresponding to the first frame of the test video

**步骤 4** 利用  $G_k$  来分析子区域的图像亮度水平,利用  $P_k$  来分析子区域的纹理复杂度,然后通过  $G_k$  和  $P_k$  选择低纹理复杂度、高平均灰度值的  $N$  个子区域作为目标区域。其理论依据是:纹理平坦的子区域能更好的体现相邻像素间的明暗变化,亮度较高的子区域能避免阴影区域的干扰,依此原则选择的目标区域可以较好反映图像的光照方向信息。具体方法如下:首先从 16 个子区域中选择  $M$  个  $P_k$  较小的子区域,再从  $M$  个子区域中选择前  $N$  个  $G_k$  较大的子区域作为目标区域,这里  $M \in [1, 16], N \leq M$ 。以图 5 中  $G_k$  和  $P_k$  分布为例,当  $M=8, N=3$  时,此时选择的目标区域为第 8、第 14 和第 15 个子区域。

**步骤 5** 使用式(5)计算各帧  $N$  个目标区域图像各自的二维光照方向矢量  $L_k^s$ ,其中  $s$  表示帧序号, $k$  表示目标区域的序号,最后得到含有  $N$  个矢量元素的二维光照方向向量  $L = \{L_k^s, k=1, 2, \dots, N\}$ 。

**步骤 6** 将步骤 3 中各帧目标区域图像对应的边缘信息百分比作为权重,对各帧二维光照方向向量  $L$  中  $N$  个矢量元素进行加权计算,得到该帧图像的光照方向矢量  $L^s$ ;加权公式如下

$$L^s = \left( \frac{1}{P_k} \times L_k^s \right) \quad (7)$$

**步骤 7** 通过计算各帧图像二维光照方向矢量的  $X$  分量和  $Y$  分量正切值所对应的角度  $\theta$  来表示该帧的二维光照方向,然后逐帧保存每帧图像的光照方向角度,得到测试视频的光照角度序列; $\theta$  的表示规则如图 6 所示,  $\theta \in [-180^\circ, 180^\circ]$ , 测试视频光照方向的角度分布如图 7 所示。

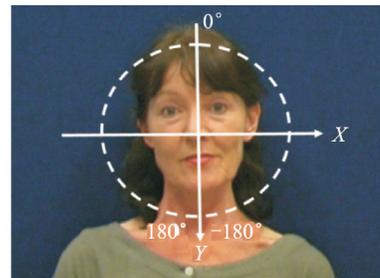


图 6 二维光照方向的角度  $\theta$  定义  
Fig.6 Definition of the angle  $\theta$  calculated by two-dimensional illumination direction

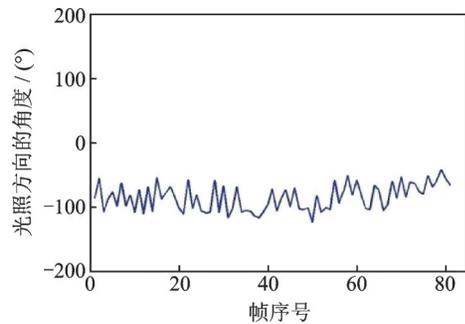


图 7 测试视频光照方向的角度分布  
Fig.7 Distribution of the angle  $\theta$  calculated by illumination direction of the test video

### 3 检测区域和目标区域数量选择

由第 2 节算法原理可知,检测区域和目标区域数量的选择是光照计算过程中的两个重要步骤,会直接影响光照方向的计算结果,因此本节通过实验对检测区域和目标区域数量的选择做进一步探究。

#### 3.1 检测区域的选择

为了测试不同大小的人脸区域对光照计算结果的影响,实验分别选择整帧区域、包含部分背景的人脸区域和人脸区域作为候选区域(分别记为类别 1、类别 2 和类别 3),然后计算真实视频和其对应换脸视频在不同候选区域中光照方向角度的分布,选择在上述两个视频中分布差异最大的候选区域作为检测区域  $A_b$ ,最后统计多段视频的  $A_b$  概率分布情况。 $A_b$  的选取可由式(8)得到

$$A_b = \operatorname{argmax}_A \left\{ D(A) = \left| \operatorname{std}_o(A) - \operatorname{std}_c(A) \right| \right\} \quad (8)$$

式中:  $\operatorname{std}_o(A)$  和  $\operatorname{std}_c(A)$  分别表示在检测区域为  $A$

时真实视频和其对应换脸视频光照方向的角度分布对应的标准差,  $D(A)$  表示  $\text{std}_o(A)$  和  $\text{std}_c(A)$  的差值绝对值。

实验通过 TIMIT 数据库中 200 对原始视频和其对应的换脸视频给出  $A_b$  的统计分布情况。结果如图 8 所示, 真实视频和换脸视频光照方向角度分布差异最大的  $A_b$  主要集中在类别 3 (即人脸区域) 处, 其原因是换脸视频仅针对人脸区域进行篡改, 使得真实视频和换脸视频的差异主要集中在人脸区域, 在背景区域则无明显差别, 因此本文算法选择视频帧图像的人脸区域作为检测区域。

### 3.2 目标区域数量的选择

首先通过一对帧长为 98 的真实视频和换脸视频给出目标区域数量为 3, 4, 5 时的光照方向结果, 这里  $M=8$ 。

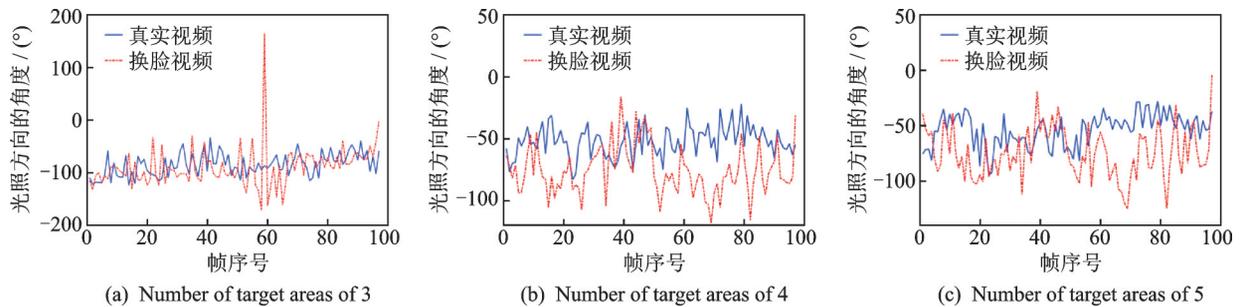


图 9 不同目标区域数量时真实视频和换脸视频光照方向的比较

Fig.9 Comparison of the illumination direction angle between the real video and its corresponding fake video with different numbers of target areas

为了选择合适的目标区域数量, 实验计算真实视频和其对应的换脸视频在目标区域数量取不同值时光照方向的角度分布, 然后选择使得真实视频和换脸视频光照方向的角度分布差异最大的目标区域数量  $N_b$ , 最后统计多段视频的  $N_b$  分布情况。  $N_b$  的选取由公式(9)得到

$$N_b = \operatorname{argmax}_N \left\{ D(N) = \left| \text{std}_o(N) - \text{std}_c(N) \right| \right\} \quad (9)$$

式中:  $\text{std}_o(N)$  和  $\text{std}_c(N)$  分别表示在目标区域数量为  $N$  时真实视频和其对应换脸视频光照方向角度分布的标准差;  $D(N)$  表示  $\text{std}_o(N)$  和  $\text{std}_c(N)$  的差值绝对值。

实验通过 TIMIT 数据库中 200 对原始视频和其对应的换脸视频给出  $N_b$  的统计分布情况。结果如图 10 所示, 使得真实视频和其对应的换脸视频光照方向分布差异最大的  $N_b$  主要集中在分布在  $N=3$  处, 因此本文算法将目标区域数量  $N$  设置为 3。

## 4 实验结果与分析

实验测试的硬件环境为 Windows 10 操作系

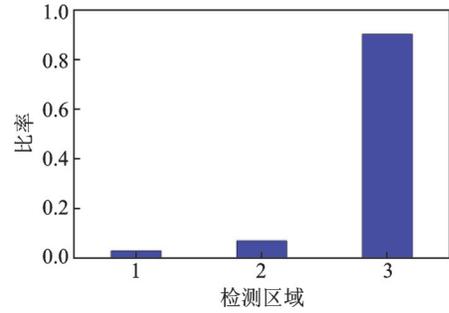


图 8 不同检测区域的概率分布

Fig.8 Probability distribution with different detection areas

结果如图 9 所示, 对于同一个测试视频, 当目标区域数量取不同值时视频光照方向角度的分布是不同的。选择合适的目标区域数量可以有效扩大真实视频和换脸视频光照方向角度分布的差异, 因此对目标区域数量的选择进行研究十分必要。

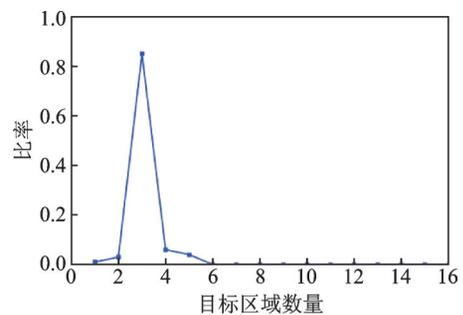


图 10 不同目标区域数量的概率分布

Fig.10 Probability distribution with different numbers of target areas

统, CPU 为 Inter(R) Core(TM) i7 - 7700, 主频 3.60 GHz, 内存 16.0 GB; 算法仿真平台为 MATLAB 2017a; 测试数据库为 TIMIT 数据库<sup>[21]</sup>和 FaceForensics<sup>++</sup> 数据库<sup>[22]</sup>, 其中 TIMIT 数据库中对每一个真实视频分别使用不同的训练模型和融合技术生成高质量 (High quality, HQ) 和低质量 (Low quality, LQ) 的两种换脸视频, 而 FaceForensics<sup>++</sup> 数据库中包含 Face2Face、Faceswap 和 Deepfake 篡改方法合成的 3 类换脸视频。

### 4.1 算法有效性验证

本节首先从 TIMIT 数据库(HQ 视频部分)和 FaceForensics<sup>++</sup>数据库(Deepfake 部分)中各选取一对真实视频和其对应的换脸视频,给出算法在个例测试对象上的实验结果。如图 11,12 所示,为了反映相邻帧光照方向角度的变化情况,实验将真假视频光照方向的角度分布进行差值处理(相邻两帧中的后一帧光照方向的角度减去前一帧光照方向的角度),结果表明两数据库中的真实视频和换脸视频光照方向的角度分布均具有明显差异,其中真实视频光照方向的角度变化较为平滑,而换脸视频光照方向的角度变化则较为剧烈。

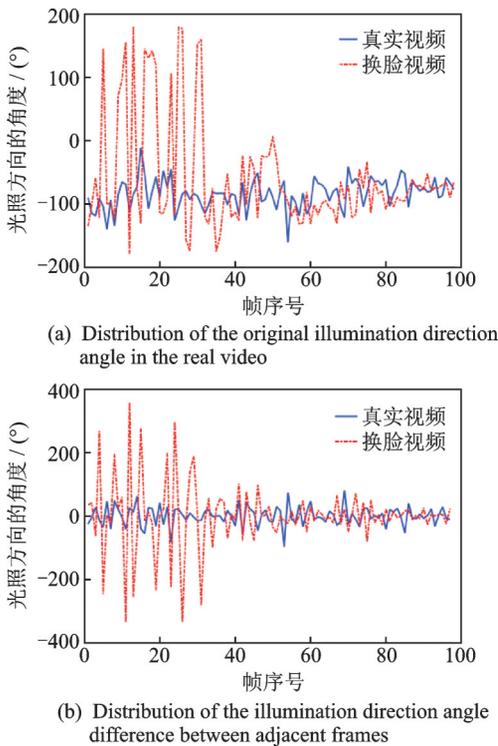


图 11 TIMIT 数据库中 faks0\_si943 真实视频和其对应换脸视频光照方向的角度分布

Fig.11 Distribution of illumination direction angle in the real video faks0\_si943 and its corresponding fake video in TIMIT database

为了进一步验证算法的有效性,实验对 TIMIT 数据库(包括 HQ 视频和 LQ 视频两部分)和 FaceForensics<sup>++</sup>数据库(Deepfake 部分)进行统计分析。首先计算数据库中每对真实视频和换脸视频光照方向角度序列的标准差,通过标准差来衡量其光照方向的角度变化是否剧烈,标准差越大,表明整段视频光照方向的角度变化越大,然后统计数据库对应的标准差分布情况。实验结果如图 13 和图 14 所示,结果表明两个数据库中真实视频和换脸视频光照方向角度的标准差分布均有明显差异,其中真实视频光照方向角度分布的标准差大多低于换脸视频光照方向角度分布的标准差,这种现象

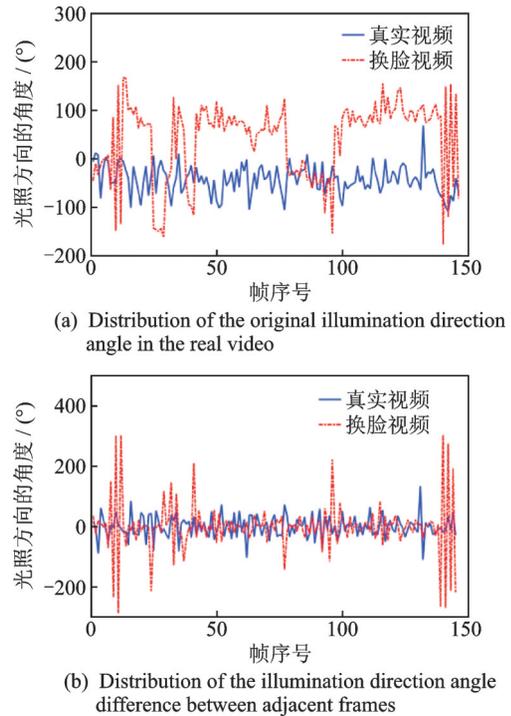


图 12 FaceForensics<sup>++</sup>数据库中 135\_880 真实视频和其对应换脸视频光照方向的角度分布

Fig.12 Distribution of illumination direction angle in the real video 135\_880 and its corresponding fake video in FaceForensics<sup>++</sup> database

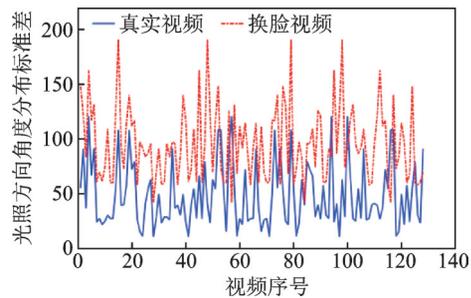


图 13 TIMIT 数据库的统计分析结果

Fig.13 Statistical analysis of TIMIT database

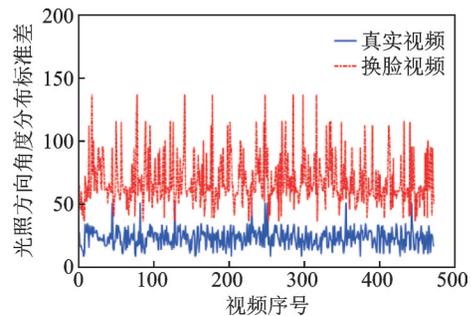


图 14 FaceForensics<sup>++</sup>数据库的统计分析结果

Fig.14 Statistical analysis of FaceForensics<sup>++</sup> database

在 FaceForensics<sup>++</sup>数据库中更加明显,由计算可得 TIMIT 数据库中真实视频和换脸视频对应标准差分布的均值分别为 49.70 和 97.14, FaceForensics<sup>++</sup>数据库中真实视频和换脸视频对应标准差

分布的均值分别为23.22和63.31。这表明所提取的光照方向特征可以较好地区分真实视频和其对应的换脸视频,证明了本文方法的有效性。

#### 4.2 算法性能分析

为了研究本文算法的性能,实验分别选择TIMIT数据库(包括HQ视频和LQ视频两部分)和FaceForensics<sup>++</sup>数据库(Deepfake部分)中真实视频和换脸视频为正样本和负样本,绘制算法的ROC(受试者工作特征)曲线。结果如图15所示,算法在FaceForensics<sup>++</sup>数据库上的检测效果要优于TIMIT数据库;且算法在TIMIT数据库中LQ视频部分的检测性能要优于HQ视频部分。其原因如下所示:

(1)FaceForensics<sup>++</sup>数据库中视频换脸效果相对TIMIT数据库较差,脸部边缘容易出现较为明显的篡改痕迹;且相比于TIMIT数据库中单一的室内拍摄环境,FaceForensics<sup>++</sup>数据库中视频拍摄场景的光照更加丰富(例如光照良好的室外和灯光明亮的室内直播间),此条件下拍摄的真实视频其光照信息更加充分,真实视频和换脸视频的光照方向一致性差异更大,因此在FaceForensics<sup>++</sup>数据库上检测效果要更好。

(2)TIMIT数据库中LQ换脸视频的视频质量更低,帧图像的纹理更模糊,无法准确反映视频的光照方向,导致换脸视频光照方向的一致性变得更差,有利于本文算法的检测,因此算法在TIMIT数据库中LQ视频部分的检测性能要优于HQ视频部分。

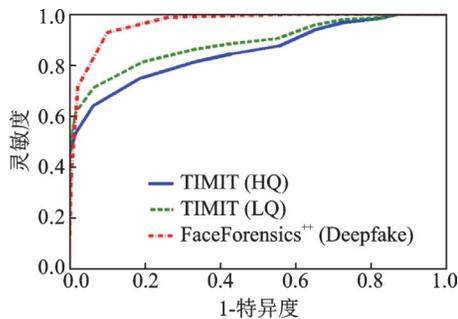


图15 算法ROC曲线

Fig.15 ROC curves of the algorithm

为了进一步验证本文算法性能,选取文献[11]、文献[12]和文献[13]作为对比算法,并使用ROC曲线下的面积(Area under the curve, AUC)作为算法性能的评价指标,分别在TIMIT数据库和FaceForensics<sup>++</sup>数据库内对不同算法进行测试,其中对比算法均以视频帧为单位进行检测,而由于本文算法充分考虑了整段视频的时域信息,因此以整段视频为单位进行检测,结果如表1所示。本文算法在2个数据库内的检测结果均高于对比

算法,其中在FaceForensics<sup>++</sup>数据库上的AUC值要远高于所对比的算法。这说明本文算法可以较好地检测真实视频和换脸视频,尤其是在视频光照环境良好的场景下,算法拥有较好的检测性能。

表1 不同算法的AUC对比

Table 1 AUC comparison of different algorithms

算法	数据库		
	FaceForensics <sup>++</sup> (Deepfake)	TIMIT (HQ)	TIMIT (LQ)
文献[11]	0.701	0.735	0.835
文献[12]	0.473	0.532	0.551
文献[13]	0.780	0.770	0.773
本文算法	0.966	0.853	0.886

由于本文算法不需要预先训练检测模型,其计算复杂度主要集中在计算待测视频帧图像的二维光照方向这一步骤,而算法所使用的Lambert光照模型是一种较为简单的单光源漫反射模型,因此本文算法具有较低的计算复杂度,实时性好。如表2所示,以帧数为100帧、分辨率为512像素×384像素的待测视频为例,本算法不需要提前训练检测模型,仅需10~15s就能完成检测,而文献[11]、文献[12]和文献[13]的方法则需要先训练好检测模型才能进行检测,且3种对比算法的检测速度均低于本文算法,故本文方法在算法实时性方面有明显优势。

表2 不同算法的实时性对比

Table 2 Realtime comparison of different algorithms

算法	测试视频 帧数/帧	测试视频 分辨率/ (像素×像素)	是否需要 训练检测 模型	检测 时长/s
文献[11]	100	512×384	是	50~60
文献[12]	100	512×384	是	15~20
文献[13]	100	512×384	是	15~20
本文算法	100	512×384	否	10~15

## 5 结 论

针对现有深度网络换脸视频篡改检测算法存在计算复杂度较高、通用性不够好以及未充分考虑视频时域特征的问题,提出一种基于光照方向一致性的换脸视频篡改检测方法。从视频成像的外部环境条件出发,利用待检测视频帧图像序列光照方向一致性进行检测,具有计算复杂度低,通用性好的特点,在公开数据库中的测试结果证明了算法的有效性。下一步的工作可考虑使用更加精确的光照模型计算视频帧图像的光照方向。

#### 参考文献:

[1] KORSHUNOVA I, SHI W, DAMBRE J, et al.

- Fast face-swap using convolutional neural networks [C]// Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 3697-3705.
- [2] BITOUK D, KUMAR N, DHILLON S, et al. Face swapping: Automatically replacing faces in photographs[J]. Proc Siggraph, 2008, 27(3): 1-8.
- [3] Impressions-video face swap App: Realistic face swap videos[EB/OL].(2020-02-15)[2020-04-05]. <https://impressions.app/>.
- [4] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 28th Conference on Neural Information Processing Systems. Montreal, Canada, 2004: 2672-2680.
- [5] 高逸飞, 胡永健, 余泽琼, 等. 5种流行假脸视频检测网络性能分析和比较[J]. 应用科学学报, 2019, 37(5): 590-608.  
GAO Yifei, HU Yongjian, YU Zeqiong, et al. Evaluation and comparison of five popular fake face detection networks[J]. Journal of Applied Sciences, 2019, 37(5): 590-608.
- [6] AFCHAR D, NOZICK V, YAMAGISHI J. Mesonet: A compact facial video forgery detection network [C]//Proceedings of IEEE International Workshop on Information Forensics and Security. Hong Kong, China: IEEE, 2018: 1-7.
- [7] BAYAR B, STAMM M C. Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11): 2691-2706.
- [8] TARIQ S, LEE S, KIM H, et al. Detecting both machine and human created fake face images in the wild [C]//Proceedings of ACM 2nd International Workshop on Multimedia Privacy and Security. Toronto, Canada: ACM, 2018: 81-87.
- [9] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 1-9.
- [10] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 1251-1258.
- [11] ZHOU P, HAN X, MORARIU V, et al. Two-Stream neural networks for tampered face detection [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, USA: IEEE, 2017: 1831-1839.
- [12] YANG X, LI Y, LYU S. Exposing deep fakes using inconsistent head poses [C]//Proceedings of International Conference on Acoustics, Speech and Signal Processing. Brighton, UK: [s.n.], 2019: 8261-8265.
- [13] MATERN F, RIESS C, STAMMINGER M. Exploiting visual artifacts to expose Deepfakes and face manipulations[C]//Proceedings of IEEE Winter Applications of Computer Vision Workshops. Hawaii, USA: IEEE, 2019: 83-92.
- [14] LI Y, CHANG M C, LYU S. In ICTU OCULI: Exposing AI created fake videos by detecting eye blinking [C]//Proceedings of IEEE International Workshop on Information Forensics and Security. Hong Kong, China: IEEE, 2018: 1-7.
- [15] PENG B, WANG W, DONG J. Optimized 3D lighting environment estimation for image forgery detection [J]. IEEE Transactions on Information Forensics and Security, 2016, 12(2): 479-494.
- [16] 牛少彰, 黄艳丽, 孙晓婷. 投影与光照方向一致性的图像篡改检测[J]. 北京邮电大学学报, 2014, 37(5): 61-65.  
NIU Shaozhang, HUANG Yanli, SUN Xiaoting. Image tampering detection by consistency of projection and lighting direction[J]. Journal of Beijing University of Posts and Telecommunications, 2014, 37(5): 61-65.
- [17] CARVALHO T, FARIA F A, PEDRINI H. Illuminant-based transformed spaces for image forensics[J]. IEEE Transactions on Information Forensics and Security, 2015, 11(4): 720-733.
- [18] DE CARVALHO T J, RIESS C, ANGELOPOULOU E, et al. Exposing digital image forgeries by illumination color classification[J]. IEEE Transactions on Information Forensics and Security, 2013, 8(7): 1182-1194.
- [19] YI J, MAO X, CHEN L, et al. Illuminant direction estimation for a single image based on local region complexity analysis and average gray value [J]. Applied Optics, 2014, 53(2): 226-236.
- [20] YANG K F, GAO S B, LI Y J. Efficient illuminant estimation for color constancy using grey pixels[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 2254-2263.
- [21] KORSHUNOV P, MARCEL S. Deepfakes: A new threat to face recognition assessment and detection [EB/OL]. (2018-12-20)[2020-04-07]. <https://arxiv.org/abs/1812.08685>.
- [22] RÖSSLER A, COZZOLINO D, VERDOLIVA L, et al. FaceForensics<sup>++</sup>: Learning to detect manipulated facial images [C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019: 1-11.