

DOI:10.19651/j.cnki.emt.1802217

# 基于 C-SIFT 特征向量图像复制粘贴篡改取证算法

都 布<sup>1</sup> 岳雅雯<sup>2</sup>

(1.大连交通大学 大连 116028; 2.上海海事大学 上海 201306)

**摘要:** 如今网络时代中充斥着大量经篡改的图像,目前检测方式如局部不变性特征描述、Harris 角点算法对复制粘贴篡改地检测准确率较低。通过对彩色图像分块、色彩空间边缘化提取、图像灰度化得到完整的灰度化局部图像。利用对不同图像块中的特征向量集提取、标记、匹配和归一化处理,在欧氏距离达到某一阈值后特征向量匹配成功,即检测到图像具有复制粘贴篡改的痕迹。最后选择 3 类不同的照片仿真测试,说明该算法可有效提升复制粘贴篡改图像的检测成功率、检测速率。

**关键词:** 图像复制粘贴篡改; C-SIFT 特征向量; 图像分块; 色彩空间边缘化提取

**中图分类号:** TN911.73    **文献标识码:** A    **国家标准学科分类代码:** 520.10

## Copy and paste tampering forensics algorithm based on C-SIFT feature vector image

Du Bu<sup>1</sup> Yue Yawen<sup>2</sup>

(1.Dalian Jiaotong University, Dalian 116028, China; 2. Shanghai Maritime University, Shanghai 201306, China)

**Abstract:** Nowadays, the network era is full of tamper-evident images. At present, the detection method such as local invariant characterization and the Harris corner algorithm has low accuracy in copying and pasting tampering. In this paper, a complete grayscale partial image is obtained by color image segmentation, color space edged extraction, and image grayscale. By extracting, marking, matching and normalizing the feature vector set in different image blocks, the feature vector is successfully matched after the Euclidean distance reaches a certain threshold, that is, the image has the trace of copying and pasting tampering. Finally, three different types of photo simulation tests are selected, which shows that the algorithm can effectively improve the detection success rate and detection rate of copying and pasting tampering images.

**Keywords:** image copy and paste tampering; C-SIFT feature vector; image segmentation; color space marginal extraction

## 0 引言

取证图像复制粘贴篡改有许多种方法,国际上数字图像盲取证上 Fridrich 等<sup>[1]</sup>提出将隐秘分析技术与盲取证结合的方法;还有学者进一步完善图像取证系统的框架,使得国外数字图像取证渐渐往数字视频取证领域过渡。国内虽起步较晚,但如计算机取证技术会议已举办多次,提出了多种解决方案,文献[2-4]提出针对一种图像局部不变性特征描述进行检测,但对彩色图像灰度信息提取地不彻底;文献[5-6]基于 SURF 算法取证,但 SURF 算法运算较慢,抗噪能力差,存在不必要角点簇等问题。基于 C-SIFT 特征向量,进行图像块之间相似度检查,提取匹配 SIFT 特征向量取证。复制粘贴篡改若仅进行复制粘贴操作则肉眼即可识别,故其处理通常伴随几何特征的改变,如添加噪点,重

压缩操作等<sup>[7]</sup>。本文提出一种将图像色彩边缘化、灰度化及分块化处理,彻底提取关键特征向量并进行匹配的方法,消除这些几何特征对检测操作的影响,准确识别图像复制粘贴篡改区域,同时提高运行效率。

## 1 图像分块化处理

提取 SIFT 特征点前需要对图像进行分块处理<sup>[8]</sup>,在彩色图像中提取 SIFT 特征点时分布通常不均匀,故特征点就需要对全局进行匹配。全局匹配不仅会消耗大量时间,同时对计算机性能要求较高。故先对图像灰度化后进行分块处理,每个图像单独进行 SIFT 特征点提取,图像块按照顺序与剩余图像块的特征向量匹配,有:

$$[m, n] = \text{size}(img) \quad (1)$$

$$nR = \text{floor}(m/N) \quad (2)$$

$$nC = \text{floor}(n/N) \quad (3)$$

式中:  $[m, n]$  分别是图像横向、纵向尺寸;  $\text{floor}$  函数用于向下取整;  $nR, nC$  分别是图像行、列块数。特征点提取完毕后,依次取按照顺序查找图像块中的特征点和剩余特征点,特征点匹配完毕后重复上述操作直到所有图像块全部完成。

## 2 彩色图像 SIFT 特征点提取

复制粘贴取证一般在图像中保持其色彩、饱和度等方面不变,SIFT 算法即尺度不变特征变换算法,它对图像的平移、缩放和旋转保持不变性的数字图像特征具有很好的取证效果<sup>[9-10]</sup>。但算法主要针对灰度图像处理,对于彩色信息处理程度有限,需要对 RGB 空间图像进行边缘提取,利用灰度共生矩阵提取不同彩色空间的边缘特征<sup>[11]</sup>,对该空间进行图像分块,不同图像块内提取足够的特征点,使用 SIFT 算法对特征点归类并进行特征匹配,如图 1 所示。

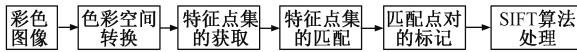


图 1 复制粘贴取证技术流程

### 2.1 色彩空间边缘提取

图像灰度化后只能保留每个像素点的亮度值,文献<sup>[12]</sup>提出了一种利用共生矩阵和自动阈值技术提取彩色图像边缘的算法。利用共生矩阵带来的统计特征构造新的图像,使用阈值技术获得最终的边缘提取。彩色图像进行灰度化,计算相邻像素点在某一个距离  $d$ , 特定方向  $\theta$  上像素分割的频率,定义为 CM,即空间信息相关法。

设  $g_{xy}$  是点  $(x, y)$  处像素值  $P_{xy}$  的强度,  $w_{xy}$  是以点  $(x, y)$  为中心,大小为  $(t \cdot t)$  的矩形窗口所包含的区域,窗口的大小影响共生矩阵的计算,计算区域内一对像素出现的频率关联值  $R$ ,则有:

$$CO(i, j, R) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \begin{cases} 1, & I(x, y) = i \\ 0, & \text{其他} \end{cases} \quad (4)$$

$$I(x' = x + d_x, y' = y + d_y) \quad (5)$$

式中:  $CO(i, j, R)$  是一对象素点  $(i, j)$  的数量,表示一个像素点灰度  $i$  与一个像素点灰度  $j$  水平相邻同时出现的频率;  $d_x, d_y$  是像素与相邻区域间的水平,垂直距离。在  $d_x$  与  $d_y$  共同的影响下,旋转操作会使得共生矩阵值发生改变,定义 4 个角度  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  计算 CM 中的特征值,然后取平均值避免共生矩阵值的较大改变。计算当前像素点  $(x, y)$  与其邻域范围内  $N(x, y)$  的加权值,得到多个边缘强度,给定每种模式系数,二维矩阵数字卷积计算,设  $LM(x, y)$  是多个边缘强度的最大值,有:

$$LM(x, y) = \max\{ES_1, ES_2, \dots, ES_n\} \quad (6)$$

给定红,绿,蓝的判别阈值  $(T_R, T_G, T_B)$ ,通过  $E_R, E_G, E_B$  函数分别对每个图像的红,绿,蓝像素进行分类,分成边缘类像素(值为 1)、非边缘类像素(值为 0)。

$$E_R(x, y) = \begin{cases} 1, & LM_R(x, y) \geq T_R \\ 0, & LM_R(x, y) < T_R \end{cases} \quad (7)$$

$$E_G(x, y) = \begin{cases} 1, & LM_G(x, y) \geq T_G \\ 0, & LM_G(x, y) < T_G \end{cases} \quad (8)$$

$$E_B(x, y) = \begin{cases} 1, & LM_B(x, y) \geq T_B \\ 0, & LM_B(x, y) < T_B \end{cases} \quad (9)$$

一般来说,如果直接使用现有的边缘检测器,经常会出现颜色边缘区域通常不连续或无法检测<sup>[13-14]</sup>。基于共生矩阵统计特征提取和阈值分析可以消除这类情况,同时又能有效消除椒盐噪声,对高斯白噪声也有一定的抑制作用<sup>[15]</sup>。

### 2.2 C-SIFT 特征点集获取

彩色图像边缘提取后,就能确定被识别图像区域,作为预先获取特征点区域<sup>[16-17]</sup>。获取 SIFT 特征点可以直接从该区域获取,大量节约了检索整张图像的时间。特征点集还包括了特征点的方向角度和尺度大小,从而形成一个完整的 SIFT 特征向量。从特征向量的尺度空间进行分析检测,灰度化图像  $I(x, y)$  分别与不同尺寸坐标系下的尺度因子  $\sigma$  下的高斯算子  $H(x, y, \sigma)$  作卷积运算,得到双正态总体变量。

$$L(x, y, \sigma) = H(x, y, \sigma) * I(x, y) \quad (10)$$

$$H(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (11)$$

得到双正态总体变量后,再用邻近的空间相关函数相减,有高斯差分变量。

$$H_1(x, y, s\sigma) = (H(x, y, k\sigma) - H(x, y, \sigma)) \quad (12)$$

$$D(x, y, \sigma) = H_1(x, y, s\sigma) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (13)$$

此方法是 LOG 算子的相近运算,将这种运算符作为检测的重要参数原因如下:

1) D 运算符计算量大幅度减少,只需要做一个减法运算,简化了算法;

2) D 运算符寻找出的特征点比较稳定,没有较大幅度上的改变,与 LOG 算子相似。

高斯差分变量构造方式如图 2 所示,图像  $I(x, y)$  (图 2(a))与 D 运算符(图 2(b))分别是两个楼层。 $I(x, y)$  的楼被分成多组,每一组又被分成多层。层与层之间差的是尺度,也就是式(12)、(13)中  $k$  取不同值带来的后果,它又叫变化尺度因子。为了在整栋楼内获取 D 运算符的极值,需要生成  $n$  个尺度间隔中  $n+3$  个高斯平滑图像,而下一组底层因子则由上一组尺度为  $2\sigma$  图像进行减少采样的方式得到,这样 D 运算符楼层就可以通过减法运算得到。

楼层中各个高斯平滑图像的因素  $\sigma$  可按照式(14)进行计算。

$$\sigma(o, s) = \sigma_0 2^{(o+s)/s} \quad (14)$$

式中:  $\sigma_0$  为最底层尺度因子;  $o$  为图像中所在组的坐标;  $s$  为图像中所在列的坐标。

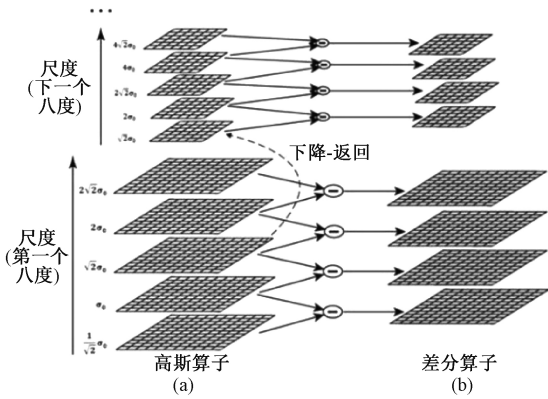


图 2 高斯差分变量构造方式

另外,空间坐标中  $x$  是组坐标  $o$  的映射,假设  $x_0$  是某一组内的空间坐标,故:

$$x = 2^o x_0, x_0 \in [0, \dots, N_0 - 1] \times [0, \dots, M_0 - 1] \quad (15)$$

式中:  $(N_0, M_0)$  为第  $o$  组中图像分辨率。楼层构造完毕后可以检测 D 运算符局部极值了。

如图 3 所示,规则是每一个像素点都必须和同一尺度中周围区域 8 个像素点以及不同尺度(一般是相邻)相同位置周围区域 18 个像素点进行对比。当且仅当被检测的点 D 运算符值大于或小于这些相比较的点后,才能确定此点为极值点。

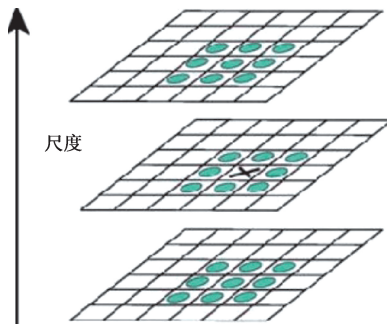


图 3 尺度楼层

拟合三维二次函数确定关键点的位置与所在的尺度空间,消除一部分减少对比度的关键点。去掉位于边缘处不稳定的点,可以提高特征匹配的稳定性。边缘检测算子如下,  $\mathbf{H}$  为 D 运算符的伴随矩阵。

$$\frac{\text{Tr}(\mathbf{H})^2}{\text{Det}(\mathbf{H})} < \frac{(r+1)^2}{r} \quad (16)$$

确定了关键点  $I(x, y)$  的位置与所在的尺度空间后,需再确定其方向,每一个图像的模值  $m(x, y)$  与方向由式(17)得到。

$$m(x, y) = \sqrt{L_1^2 + L_2^2} = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (17)$$

$$\theta(x, y) = \tan^{-1} \left( \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right)$$

由此可得 SIFT 特征点位置、尺度和方向特征。

### 2.3 特征点集的匹配与标记

设特征点此时的方向为  $\theta$ , 因生成的特征向量旋转会差出  $2\theta$ , 故之前先旋转领域  $-\theta$  角度, 然后对领域进行分块操作接着让各块统计做一个梯度直方图, 这里可以使用特征点方向确定时的办法。这样得到的多维特征向量消除了缩放、旋转等操作带来的干扰。最后还要对特征向量进行归一化处理, 对比度发生变化表现为各个像素模值和方向呈线性变化关系<sup>[18]</sup>, 光照会使得每个像素点都增加一个常数, 这也就说明了它们之间的差值是不会发生变化的, 但如果非线性光照可能会使常数增加的不一致, 特征点会发生影响。相似性匹配通过计算两个特征点之间的欧氏距离进行相似度的度量。

不管是一幅图片中的两个部分, 还是两幅图片中的某一部分, 因隐藏遮蔽等因素不能很好地进行匹配, 更有可能出现误配的情况。对此可以在匹配时拿出与该特征(关键点)欧氏距离最相近和次相近的两个关键点, 用最近的关键点欧氏距离除以次最近的欧式距离, 选定一个阈值, 低于此阈值, 该对关键特征(关键点)匹配成功<sup>[19]</sup>。

$$\frac{\text{Distance}(1)}{\text{Distance}(2)} < r \quad (18)$$

通过式(18)可知, 如果该阈值数值较低(即选定时更加地严格), 则稳定性会增强, 但会增加丢失匹配对的概率; 如果阈值过高, 那么会出现误判的情况, 需要依实际情况进行选定。

### 3 仿真测试

本文在 MATLAB 仿真平台进行测试, 为了确保算法的实用性和广泛性, 本文选取了 3 种不同类型照片的代表。来自相机并用 PS 复制粘贴篡改如图 4 所示; 图 5 所示为网上摘取的新闻图片; 图 6 所示为手机图片并用 PS 复制粘贴篡改, 实际测试结果如下。



图 4 小船取证图

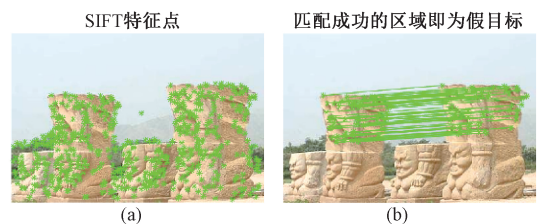


图 5 雕塑取证图

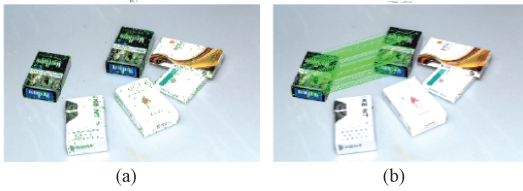


图 6 香烟取证图

检测时所使用的 PC 配置环境是英特尔 i5-4 300 u, 4 G 内存, 英特尔集成显卡, 3 种来源的图像分别选取同一类型的图像 200 张, 表 1 和 2 说明了使用图像分块和彩色边缘化处理后基于 SIFT 特征点取证的成功率和平均识别速率。

表 1 SIFT 特征点取证成功率 (张)

分块	类型 1	类型 2	类型 3
分块前	163/200	174/200	194/200
分块后	170/200	188/200	195/200

表 2 SIFT 特征点取证识别速率 (s)

分块	类型 1	类型 2	类型 3
分块前	5.334	45.93	3.59
分块后	3.9	32.77	2.22

表 1 说明图像在经过分块和边缘化处理后识别成功率有了一定地提升, 分块后的图像在边缘化提取上失去的边缘信息较未分块少, 特征向量能尽量避免干扰, 准确匹配。类型 1、2 图像选取的色彩较为简单且篡改操作只有复制粘贴, 识别提升较大, 类型 3 图像颜色较为复杂对识别造成了一定的影响, 整体上识别效果较好。表 2 说明图像分块边缘化处理后取证识别速率大幅提高, 主要原因是边缘化处理在图像灰度化时减少对冗余区域地检测, 特征向量在不同图像块之间比对次数降低。表 1 和 2 整体说明基于 C-SIFT 特征向量的方式能有效检测图像识别篡改区域。表 3 同与文献[2-6]相比, 本文所使用的图像分块后基于 C-SIFT 特征取证的算法有着更好的适应性和准确性, 识别成功率优于其他算法, 在实践应用上识别准确快速。

表 3 C-SIFT、SURF、HARRIS 算法比较

类型	C-SIFT	SURF	HARRIS
成功率/%	95.63	82.77	86.53

受限于图像拍摄模式, 场地环境的多样性, 识别会出现一定地失误, 以后的研究可以自适应调整分块大小和特征向量识别, 进一步提升识别的准确性。

## 4 结 论

本文基于 C-SIFT 特征向量对数字图像进行盲取证,

图像经过色彩空间边缘化提取、灰度化后较之前的图像在后续检测成功率上有明显提高。仿真测试说明图像经分块后识别准确程度提高 7% 左右, 运算时间被降低大约 5 s, 充分解决了目前数字图像盲取证实际使用中存在的识别成功率低, 识别速度慢的缺点。在和其余算法对比中基于 C-SIFT 算法成功率优于其余文献给出的算法, 但对一些存在其他篡改操作的图像识别成功率有待加强, 可以在边缘化处理和图像分块上做进一步研究和分析。

## 参 考 文 献

- [1] FRIDRICH J, SOUKAL D, LUKAS J. Detection of copy-move forgery in digital images[C]. Proceedings of Digital Forensic Research workshop, 2003.
- [2] 朱仁欢. 局部不变特征描述算法的研究及其应用[D]. 合肥: 安徽大学, 2015.
- [3] 王永明, 王贵锦. 图像局部不变性特征与描述[M]. 北京: 国防工业出版社, 2010.
- [4] 许允喜, 陈方. 局部图像描述符最新研究进展[J]. 中国图象图形学报, 2015, 20(9): 1133-1150.
- [5] 李岩, 刘念, 张斌, 等. 图像镜像复制粘贴篡改检测中的 FI-SURF 算法[J]. 通信学报, 2015, 36(5): 58-69.
- [6] 高宝建, 金媛媛, 柏潇. 基于 SURF 的图像复制篡改检测[C]. Proceedings of 2010 Asia-Pacific Conference on Information Network and Digital Content Security (2010APCID), 2010: 75-79.
- [7] 池静, 张婷, 杨振宇. 基于 SIFT 和区域生长的图像复制粘贴盲检测算法[J]. 科学技术与工程, 2013, 13(4): 917-920, 925.
- [8] 杜媛. 基于图像分割与多特征模型的车牌识别算法[J]. 国外电子测量技术, 2018, 37(8): 20-24.
- [9] 曹雁军. 数字图像被动取证与反取证技术研究[D]. 天津: 南开大学, 2013.
- [10] 骆伟祺, 黄继武, 丘国平. 鲁棒的区域复制图像篡改检测技术[J]. 计算机学报, 2007, 30(11): 1998-2007.
- [11] 闵永智, 岳彪, 马宏锋, 等. 基于图像灰度梯度特征的钢轨表面缺陷检测[J]. 仪器仪表学报, 2018, 39(4): 220-229.
- [12] 陈小芬, 王跃存. 基于 Kirsch 理论的彩色图像边缘检测[J]. 仪器仪表用户, 2008, 15(5): 111-112.
- [13] WANDJI N D, SUN X, KUE M F. Detection of copy-move forgery in digital images based on DCT[J]. Computer Science, 2013, 111(1): 148-65.
- [14] AMERINI I, BALLAN L, CALDELLI R, et al. A SIFT-based forensic method for copy-move attack detection and transformation recovery [J]. IEEE Transactions on Information Forensics & Security, 2011, 6(3): 1099-1110.
- [15] 王威. 数字图像篡改盲取证技术研究[D]. 宁波: 宁波大学, 2011.

- [16] POPESCU A C, FARID H. Exposing digital forgeries in color filter array interpolated images [J]. IEEE Transactions on Signal Processing, 2005, 53 (10): 3948-3959.
- [17] 刘丽娟,林小竹.基于背景噪声的图像盲篡改检测[J]. 计算机科学,2014,41(S2):136-138,149.
- [18] 王森,伍星,张印辉,等.基于多尺度降采样规范化割的图像裂纹检测[J]. 仪器仪表学报,2017,38(11): 2788-2796.

- [19] 王鑫,轩波,彭思龙.数字彩色图像拷贝-变换-移动篡改检测[J]. 计算机工程,2009,35(16):20-22.

### 作者简介

都布,硕士研究生,主要研究方向为高速列车无线信道与通信信息技术。

E-mail:dubu1994@tju.edu.cn

岳雅雯,硕士研究生,主要研究方向为多电机同步控制、滑模控制。

E-mail:yyw0127dhp@qq.com

## 泰克携手西安电子科技大学微电子学院共建微电子实践育人平台

近日全球领先的测量解决方案提供商泰克科技(Tektronix)携手西安电子科技大学微电子学院共建示范性微电子实践育人平台,揭幕仪式在西安电子科技大学微电子学院举行。泰克科技大中华区和东南亚区总裁洪斌顺先生与中科院院士、西安电子科技大学郝跃教授共同为平台揭幕。同时,洪斌顺先生代表泰克公司向郝跃院士颁发了泰克公司全球半导体顾问证书。

同时参加揭幕的还有西电微电子学院党委书记黄伟、微电子学院副院长马晓华、宽禁带半导体重点实验室副主任马佩军,以及微电子学院教师代表。郝院士为平台的成立致辞,郝院士提到中国的半导体与集成电路产业取得了长足的进步,同时也指出中国的半导体发展缺乏“从 0 到 1”的原始创新,而创新离不开人才的培养,西电-泰克“微电子实践育人平台”的建立是一次创新微电子人才培养模式的探索。郝院士提出平台的建立将实现“三个有利于”,有利于微电子专业的人才培养,有利于扩大泰克品牌在中国教育领域尤其是微电子方向的影响力,有利于双方进一步在教育平台建设、人才培养、产学研协同创新方面进行合作。同时也对平台后续的发展提出殷切希望,双方表示将继续加强合作,将平台越做越好。

洪斌顺先生表示,今天的仪式只是一个开始,未来泰克将投入更多的资源在微电子的创新技术领域,在与西电的实践育人平台中分享更多泰克最新技术发展,为实现中国

微电子产业的创新升级注入源源不断的动力。西安电子科技大学是教育部直属的全国重点大学,是国家“优势学科创新平台”项目和“211 工程”项目重点建设高校之一、国家双创示范基地之一、首批 35 所示范性软件学院、首批 9 所示范性微电子学院、首批 9 所获批设立集成电路人才培养基地和首批一流网络安全学院建设示范项目的高校之一。

随着西电-Tektronix 微电子实践育人平台的建立,双方将开展微电子学科建设与专业领域研究合作,达成建设为国家输送集成电路设计、制造、封装测试及其装备、材料等方向的工程型人才的培养基地的目标。针对微电子专业基础课程与实验课程的实践性强、教学内容更新快、条件建设昂贵等特点,双方决定进行创新模式的校企合作,以便更好地提供优质的教学实践条件和实践资源,有效提高学生的实验动手能力与工程实践技能。

这个前瞻性的微电子实践育人平台,正是产学研融合协同育人的示范性平台。通过双方在微电子领域产学研合作协同育人、科研创新和产业转化,以及带动学科的整体发展和影响力,联合创新集成电路人才培养机制,共同制定培养目标、共同设计课程、共同开发教材、共建教学团队、共建实习实践实训平台,在全国相关高校及半导体行业内起到示范和引领作用,打造微电子领域具有前瞻性的产学研融合协同育人的示范性平台。更多内容请访问 <https://www.tek.com.cn>。