计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.250900081

# 基于量子安全和脆弱水印的图像篡改检测与自恢复算法

陈鸿祥[1] 陈 果[1] 张 辉[1] 吴美琪[1] 丁萁琦[1] 罗 合[1] 王 昊[1] 谷林明[1] 罗惠恒[2] 王景晗[2]
1 中国长江电力股份有限公司 武汉 430014
2 中国三峡武汉科创园 武汉 430000

**摘 要** 在电力系统的远程监控与无人值守巡检应用中,图像已成为记录与传输缺陷信息的重要媒介。然而,经由公开信道传输的图像数据极易受到恶意篡改与伪造,严重威胁系统安全与故障响应效率。为此,文中提出了一种适用于电力通信场景的图像篡改检测与自恢复算法。该算法基于量子随机数生成器构造完美哈希,用于生成具备高随机性与不可预测性的认证模型,以提高其抗篡改能力;同时结合图像块级匹配策略和SPIHT编码算法生成认证与恢复数据;然后,将这些数据作为脆弱水印嵌入到原始的通信图像中,以实现篡改区域的精确定位和自恢复,其中嵌入密钥采用量子密钥分发协议进行安全共享,可有效防止关键参数在传输过程中的泄露或篡改。在标准图像数据集上进行实验,结果表明所提方法在篡改检测精度、自恢复效果、安全性、嵌入容量与图像质量等方面均优于现有方案,适用于电力图像传输的完整性保护与可信认证需求。

**关键词**:篡改检测;图像恢复;脆弱水印;量子安全;完美哈希
**中图分类号** TP391

# Image Tampering Detection and Self-recovery Algorithm Based on Quantum-secure and Fragile Watermarking

CHEN Hongxiang[1],CHEN Guo[1],ZHANG Hui[1],WU Meiqi[1],DING Qiqi[1],LUO He[1],WANG Hao[1],GU Linming[1],LUO Huiheng[2] and WANG Jinghan[2]
1 China Yangtze Power Co.,Ltd.,Wuhan 430014,China
2 China Three Gorges Wuhan Science and Technology Innovation Park,Wuhan 430000,China

**Abstract** In the context of remote monitoring and unattended inspection in power systems,images have become a crucial medium for recording and transmitting defect information. However,when transmitted over public channels,these images are highly susceptible to malicious tampering and forgery,posing serious threats to system security and fault response efficiency. To address this issue,this paper proposes an image tamper detection and self-recovery method tailored for power communication scenarios. The proposed approach constructs a perfect hashing-based authentication model using a quantum random number generator,ensuring high randomness and unpredictability to enhance resistance against tampering. It further integrates a block-level image matching strategy and a SPIHT encoding algorithm to generate both authentication and recovery data. These data are embedded into the original communication image as fragile watermarks,enabling precise localization and recovery of tampered regions. The embedding key is securely shared via a quantum key distribution protocol,effectively preventing key leakage or manipulation during transmission. Experimental results on standard image datasets demonstrate that the proposed method outperforms existing schemes in terms of tamper detection accuracy,recovery performance,security,embedding capacity,and visual quality,making it well-suited for integrity protection and trusted authentication of power image transmission.

**Keywords** Tamper detection,Image recovery,Fragile watermarking,Quantum security,Perfect hash

## 1 引言

随着电力系统规模的不断扩大以及智能电网技术的广泛部署,输电线路的运行安全与故障快速响应能力正面临着前所未有的挑战。传统的人工巡检方式已难以满足大范围、长距离、高复杂度输电环境下的运维效率要求,远程视频监控与图像采集技术逐渐成为输电线路状态感知与缺陷检测的重要手段。在此背景下,图像因其具有高信息冗余性、可视化直观、便于处理与传输等特点,成为了承载电力设备缺陷信息的核心媒介,被广泛用于辅助决策与运维调度。然而,由于通信链路多采用开放信道,图像数据在传输过程中极易受到诸如篡改、伪造与重放等恶意攻击,从而影响接收端判断故障区域的准确性,威胁电力系统的安全性与可靠性。因此,构建具备篡改检测与内容恢复能力的图像认证机制,已成为电力系统通信安全中亟待解决的关键问题。

近年来,图像篡改检测与自恢复技术已成为多媒体安全领域的重要研究方向。相关技术可大致分为两类:被动认证方法与主动认证方法[1]。被动认证方法[2-4]通常依赖于图像内部的统计特征或几何信息来实现篡改检测,代表技术包括频域特征分析、深度学习异常检测等,但这类方法通常难以实现篡改区域的精确恢复,且在对抗恶意攻击方面的鲁棒性不足。相比之下,主动认证方法[5-7]则通过将认证数据嵌入到原

基金项目:中国长江电力股份有限公司科研项目(1623020020)
This work was supported by the Scientific Research Project of China Yangtze Power Co.,Ltd.(1623020020).
通信作者:陈鸿祥(chen_hongxiang@ctg.com.cn)

始图像中，在保证图像视觉质量的前提下，实现对篡改区域的检测、定位以及内容恢复。其中，脆弱水印(Fragile Watermarking)因其对图像微小修改的高度敏感性，已成为实现高精度图像篡改检测与自恢复系统的重要方法[8]，尤其适用于对图像真实性与完整性具有严格要求的场景，如电力系统缺陷图像的认证与保护。Van 等[9]最早提出将数字图像的 8 比特灰度值压缩至 7 比特，并利用腾出的比特位嵌入水印，实现图像内容的嵌入式认证。随后，Lin 等[10]设计了分层脆弱水印结构，通过图像分块与平均灰度的奇偶校验构建三层检测机制，显著提升了篡改检测率，但在面对彼此映射的图像块被同时攻击时，恢复质量会显著下降。Lee 等[11]提出了双重水印机制，该机制在每个图像块中嵌入其自身认证信息与另一图像块的恢复数据，实现双向验证与冗余恢复。Qin 等[12]提出重叠嵌入策略，实现对篡改区域的精确定位与恢复。Sarreshtedari 等[13]利用图像编码方法提取图像多层系数作为恢复依据，并结合 Reed-Solomon 信道解码提高图像重建质量，但仍难以抵御 copy-move 类攻击。Bolourian 等[14]进一步引入多组摘要结构、镜像备份与邻块验证机制，有效增强了对复杂篡改行为的响应能力。为提升局部适应性，Sreenivas 等[15]将图像划分为 2×2 子块并嵌入双组认证与恢复信息，实现图像块的随机定位与内容恢复，但在应对有遮挡区域或结构复制攻击方面仍存在限制。此外，一些方法尝试通过引入编码理论与混沌映射来增强安全性。Prasad 等[16]结合汉明编码与 Logistic 映射构造认证序列，提高了水印嵌入过程的不可预测性与抗篡改性。Liu 等[17]则设计自适应水印嵌入策略，依据图像块的量化特征与像素对角映射生成认证信息，并实现基于位置扰动的非对称嵌入，部分提升了安全性与检测能力，但仍受限于量化误差导致的误判问题。综上所述，已有的方法虽在一定程度上提升了篡改检测与恢复性能，但在面向高容量嵌入、安全密钥管理及应对复杂攻击模型等方面仍存在鲁棒性弱、自恢复精度有限、认证机制易被预测等问题，难以满足电力通信图像在实际应用中对安全性与完整性的高要求。

与此同时，随着量子通信技术的快速发展，基于量子密码的安全机制为图像篡改检测与自恢复系统提供了可靠支撑。量子密码技术是一种基于量子力学基本原理(如测不准原理与不可克隆定理)构建的信息安全机制，具备对量子计算攻击的天然抵抗能力。量子密钥分发(Quantum Key Distribution，QKD)和量子随机数生成器(Quantum Random Number Generator，QRNG)作为量子保密通信体系中的核心技术，在保障信息传输的机密性、完整性和抗攻击性方面展现出了显著优势。QKD[18]可实现通信双方在不依赖复杂数学难题的前提下，安全生成并共享加密密钥，打破了传统密码体制在后量子时代所面临的安全瓶颈。QRNG 是基于量子物理过程生成不可预测的高熵随机序列，能够提供远高于传统伪随机数发生器的真随机性保障。相比传统伪随机数生成器依赖确定性算法与种子，QRNG[19]直接基于量子物理过程生成具有理论不可预测性和高度熵值的真随机数。其输出序列在统计特性、分布均衡性和可验证性等方面均优于经典方案，可显著增强密钥生成过程的安全性，并有效抵御统计分析类攻击。在图像篡改检测等场景中，QRNG 可用于构造高安全性的认证索引，有效抵御基于统计特征的推测性攻击。

针对上述问题，本文提出了一种适用于电力通信场景的图像篡改检测与自恢复算法。该算法基于 QRNG 构造完美哈希，用于生成具备高随机性与不可预测性的认证模型，以提高其抗篡改能力；同时结合图像块级匹配策略和 SPIHT 编码算法生成认证与恢复数据；然后，将所生成的这些数据作为脆弱水印嵌入到原始图像中，以实现篡改区域的精确定位和自恢复，其中嵌入密钥采用 QKD 协议进行安全共享，可有效防止关键参数在传输过程中的泄露或篡改。实验结果表明，所提方法在篡改检测精度、自恢复效果、安全性、嵌入容量与图像质量等方面均优于现有方案，适用于电力图像传输的完整性保护与可信认证需求。

## 2 理论基础

### 2.1 完美哈希

哈希函数(Hash Function)是一类将任意长度的输入数据压缩映射为固定长度数据摘要的映射函数，即实现关键字与其对应散列地址之间的映射，被广泛应用于关键字索引、数据认证与加密等领域。在图像认证场景中，其可用于实现图像块特征(作为关键字)与其对应散列地址之间的快速匹配。然而，传统哈希函数可能存在碰撞问题，即多个不同关键字被映射至相同的散列地址，从而降低认证的准确性与系统的整体效率。为解决上述问题，研究者提出了完美哈希函数(Perfect Hash Function)[20]，该函数可实现关键字空间到散列地址空间之间的无碰撞映射，在理论上为认证机制提供理想支撑。

Du 等[21]提出了一种基于重哈希(Rehashing)模型的完美哈希函数构造方法，如图 1 所示。假设关键字集合为 $K_1$，$K_2$，…，$K_n$，通过某单一哈希函数 $h_k$ 将其映射至大小为 $m$ 的地址空间 $A_1$，$A_2$，…，$A_m$，其中 $h_k$ 随机选自映射函数集合 $F_{n\times m}$。由于单一哈希函数成功构造无碰撞映射的概率极低，尤其在 $n\approx m$ 时更易发生冲突。其成功映射的概率的计算式如下：

$$P_i(m,n)=\frac{n!\binom{m}{i}\sum_{r=0}^{n-i}(-1)^r\binom{m-i}{r}\frac{(m-r-i)^{n-r-i}}{(n-r-i)!}}{m^n} \tag{1}$$

其中，$P_i(m,n)$ 表示在地址空间中有 $i(0\leqslant i\leqslant\min(m,n))$ 个散列值被唯一映射的概率。计算结果显示，当 $i\geqslant 0.8n$ 且 $n=m=10,20,30$ 时，$P_i$ 仅分别为 1.6700%，0.0205% 和 0.0003%，难以满足实际认证需求。为克服上述限制，Du 等[21]提出了一种重哈希模型：通过引入多个随机哈希函数 $h_1$，$h_2$，…，$h_k$ 对关键字进行多轮映射，从而显著降低碰撞概率。每个关键字在映射过程中所使用的哈希函数编号被记录于哈希指示表(Hash Indicator Table，HIT)，从而实现哈希路径的可控性与认证过程的可追溯性。该模型下唯一映射的概率可表示为：

$$P_i^k(m,n)=\sum_{r=0}^{i}P_r^{k-1}(m,n)\times\left(\frac{r}{m}\right)^{n-r}\sum_{s=0}^{n-r}\binom{n-r}{s}\frac{e_{i-r}(s,m-r)}{r^s} \tag{2}$$

实验结果表明，当采用 7 个哈希函数构建重哈希模型(即 $k=7$)时，参数 $i$ 的期望值分别为 8.8，17.5 和 26.1，在 $i\geqslant 0.8n$ 且 $n=m=10,20,30$ 的条件下，映射成果的概率提升至 96.4%，97.2% 和 97.9%，显著优于传统单一哈希方法。这表明重哈希策略能够显著提升哈希映射的去碰撞能力，趋近于完美哈希效果。基于上述理论基础，本文设计了一种结合完美哈希的图像块级匹配策略，通过构造块级 HIT 位信息实现对图像块数据的标识与认证。该策略在保证重哈希模型无碰撞映射的前提下，显著提升了图像篡改区域的精确定位能力，为后续的篡改检测与自恢复操作提供了高效、可靠的技术支撑。

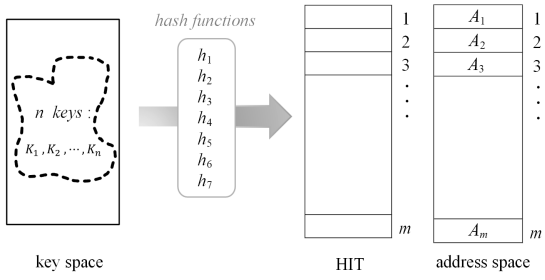



图 1　基于重哈希模型的完美哈希

Fig. 1　Rehashing model based perfect hash

### 2.2　多级树集合分裂 SPIHT

Shapiro 最早提出了基于离散小波变换(Discrete Wavelet Transform，DWT)的嵌入式零树小波(Embedded Zero-tree Wavelet，EZW)图像编码算法[22]，为后续小波域图像压缩方法奠定了理论基础。在此基础上，Said 和 Pearlman 进一步设计了一种改进型算法——多级树集合分裂(Set Partitioning in Hierarchical Trees，SPIHT)图像编码算法[23]，该算法在保持高压缩效率的同时兼具良好的嵌入特性与渐进式传输能力，成为图像重建信息提取的重要工具。其核心思想是利用小波系数在空间域和尺度域中的相关性，对图像信号的显著系数集合进行层级分裂与优先传输，以实现更高效的数据压缩与渐进式重构。由于小波变换后图像能量大多集中于低频区域，SPIHT 算法优先编码幅值较大的小波系数，并可在任意给定的码率或失真约束下截断编码过程，从而在有限比特预算下获得更优的图像质量。SPIHT 引入了空间方向树(Spatial Orientation Tree，SOT)结构将小波域中不同尺度的系数组织为具有父子层级关系的树状结构，其树结构如图 2 所示。

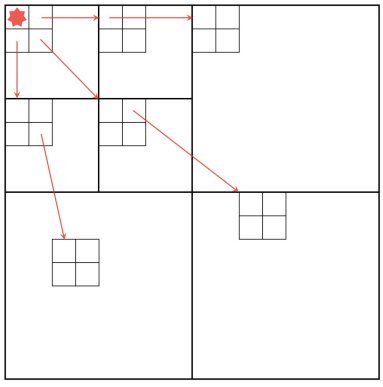


图 2　空间方向树结构图

Fig. 2　Structure diagram of spatial direction tree

在编码过程中，算法将所有系数集合划分为 3 类列表，即重要系数表(List of Significant Pixels，LSP)、不重要系数表(List of Insignificant Pixels，LIP)和不重要集合表(List of Insignificant Sets，LIS)，并基于当前阈值对各集合进行逐层判断与分裂。每一轮迭代中，算法优先输出当前最显著的系数信息，并对尚未编码的系数集合进行递归分解，从而实现嵌入式、渐进式的比特流输出。SPIHT 算法中小波系数重要性的判定公式如下：

$$S_n(X)=\begin{cases}1, & \max\left\{\left|\begin{matrix}c(i,j)\\(i,j)\in X\end{matrix}\right|\right\}\geqslant 2^n\\0, & \text{其他}\end{cases} \tag{3}$$

其中，$X$ 表示小波系数的坐标集合，$c(i,j)$ 为位置 $(i,j)$ 处的小波系数。为实现高效编码，SPIHT 定义了四类集合：所有根节点的坐标集合 $H(i,j)$；位置 $(i,j)$ 处小波系数的直接后代集合 $O(i,j)$，其元素数目为 4 或 0；位置 $(i,j)$ 处小波系数的全部后代集合 $D(i,j)$；位置 $(i,j)$ 处小波系数的所有非直接后代集合 $L(i,j)=D(i,j)-O(i,j)$。SPIHT 的编码过程主要包括初始化、排序扫描与精细扫描 3 个阶段。

1)初始化：算法首先设定阈值序列 $\{T_0,T_1,\cdots,T_{N-1}\}$，用于判定小波系数在各层级下的重要性。初始阈值定义为 $T=2^n$，其中 $n=\lfloor\log_2\max\{|c(i,j)|\}\rfloor$，$\lfloor\cdot\rfloor$ 表示向下取整。将各列表初始化为：

$$\begin{cases}LSP=\varnothing\\LIP=\{(i,j)\,|\,(i,j)\in H\}\\LIS=\{D(i,j)\,|\,(i,j)\in H\}\end{cases} \tag{4}$$

2)排序扫描：按照 EZW 零树[22]的 Morton 扫描顺序(即“Z”字形扫描)依次检查 LIP 中的每个系数。若 $S_n(i,j)=1$，则输出 1 并附加符号位(若 $c(i,j)\geqslant 0$，符号位为 1；若 $c(i,j)<0$，符号位为 0)，同时将该坐标移入 LSP 末尾；否则输出 0，且不移动其位置。对 LIS 中的元素进行如下处理：若元素属于 $D$ 集合，则依次检查其 4 个直接后代，若为重要系数则输出 1 并移入 LSP；若为非重要系数则输出 0 并移入 LIP；随后将该元素从 LIS 删除。若元素属于 $L$ 集合，则需判断该集合是否为重要集合。若为重要集合，则输出 1，同时删除该元素，并将集合中的 4 个坐标加入 LIS；若为非重要集合，则输出 0，并保留该元素在 LIS 中。

3)精细扫描：依次检查 LIP 中的各元素(不包括本次扫描中新加入的元素)，并输出对应小波系数的第 $n$ 位高比特。若 $n=0$，则编码过程终止；否则令 $n=n-1$，并返回执行排序扫描与精细扫描的循环，直至结束。

在图像恢复过程中，SPIHT 的嵌入式编码机制具备显著优势。其通过对小波系数的重要性排序并采用逐次逼近的量化方法生成的比特流，不仅压缩率高，而且利于逐级还原图像内容。然而，若嵌入区域遭受恶意篡改，SPIHT 所依赖的小波系数结构将遭到不可逆破坏，严重影响解码效果与恢复图像的视觉质量，特别是在大范围篡改场景下，传统自恢复技术难以维持有效还原能力。为解决上述问题，本文提出将 SPIHT 编码与基于完美哈希的图像块匹配策略相结合，构建鲁棒性更高的图像自恢复机制。具体而言，编码过程中将原始图像以 8 bpp 进行输入，并对 SPIHT 输出流按照 1 bpp 的速率进行截断，以控制嵌入容量。同时，结合自恢复位与映射恢复位构成完整的恢复数据，提升系统对篡改区域的鲁棒性与重建精度，从而在维持图像质量的基础上显著增强篡改恢复能力。

## 3　图像篡改检测与自恢复机制

针对现有方法在篡改检测精度、图像恢复质量、水印嵌入容量、安全性等方面存在的局限，本文提出了一种基于量子安全和脆弱水印的图像篡改与自恢复机制。该机制首先利用 QRNG 产生高熵随机数，用于构建具备高度不可预测性的完美哈希模型。基于该模型设计了图像块级匹配策略并结合 SPIHT 图像编码算法，生成认证数据与恢复数据。接着，设计了基于层级扩展的脆弱水印算法，将上述的认证数据与恢

复数据嵌入到原始的电力通信图像中，以实现篡改区域的精确定位和自恢复，其中嵌入密钥采用 QKD 协议进行安全共享，可有效防止关键参数在传输过程中的泄露或篡改，增加该机制在开放信道下的密钥保密性与抗攻击能力。

### 3.1 融合完美哈希和 SPIHT 的图像块级匹配算法

图像篡改检测与自恢复机制的核心在于生成能够准确表征原始图像特征的认证与恢复数据，并将其嵌入至原始图像中以实现篡改区域检测与内容自恢复。由于图像中每个像素位置具有唯一性，本算法将 $h\times w$ 大小的原始图像划分成若干个 2×2 大小的像素块，并将每一图像块的空间位置序号作为哈希函数的关键字，利用由 QRNG 产生的种子 $r$ 构造 7 个独立的随机哈希函数 $h_1, h_2, \cdots, h_7$。相比传统伪随机数生成器依赖确定性算法与种子，QRNG 直接基于量子物理过程生成具有理论不可预测性和高度熵值的真随机数。其输出序列在统计特性、分布均衡性和可验证性等方面均优于传统方案，可显著增强密钥生成过程的安全性，并有效抵御统计分析类攻击。在本算法中，关键字空间 $j$ 和散列地址空间 $h_k(j)$ 的大小均为 $h\times w/4$，$h_k(j)$ 表示图像块 $B_j$ 的序列号 $j$ 作为关键字经随机哈希函数 $h_k$ 映射得到的散列值。为支持高效映射与冲突检测，除了 HIT 外，本算法还设计了以下结构：散列地址表(Hash Address Table，HAT)用于记录每个关键字所对应的散列地址；逆散列地址表(Inverse Hash Address Table，IHAT)用于记录每个散列地址所对应的原关键字，以实现快速逆查；占位表(Occupied Table，OT)用于辅助识别散列地址的占用状态，提高数据查找效率。

在此基础上，本文设计了完整的水印信息结构，将认证数据 $D_A$ 与恢复数据 $D_R$ 共同嵌入到原始图像中，以实现图像篡改区域的精确检测与自恢复。其中，认证数据 $D_A$ 由 HIT 位与其对应的奇偶校验位组成，用于提高篡改检测的鲁棒性与准确性；恢复数据 $D_R$ 由图像块 $B_j$ 自身的恢复位 $D_{RS}$ 和及其逆映射图像块 $B_{IHAT(j)}$ 的恢复位 $D_{RM}$ 组成，两者结合能够增强面对大范围篡改攻击下的图像恢复能力。算法的流程如算法 1 所示。

**算法 1** 图像块级匹配算法

输入：量子随机数种子 r，h×w 原始图像 I

输出：认证数据 $D_A$ 与恢复数据 $D_R$

步骤 1 预处理与初始化

将原始图像 I 划分为不重叠的 2×2 大小的图像块 $B_j$，其中总图像块数为 h×w/4。利用量子随机数种子 r 构造 7 个独立的随机哈希函数 $h_1, h_2, \cdots, h_7$，创建长度为 h×w/4 的 HIT，HAT，IHAT 和 OT，并将表内的值全部初始化为 0。

步骤 2 构建哈希映射关系

对于每个图像块序列号 j=1，2，…，h×w/4，执行如下操作：从 $h_1$ 至 $h_7$ 依次尝试计算散列地址，若当前哈希函数 $h_k$ 计算出的地址 $h_k(j)$ 未被占用(即 $OT(h_k(j))=0$)，则记录 $HIT(j)=k$，$HAT(j)=h_k(j)$，且将 $OT(h_k(j))$ 置为 1 来标记该散列地址已被占用；若当前哈希函数 $h_k$ 计算出的地址 $h_k(j)$ 冲突，则继续尝试下一个哈希函数。对于上述过程中仍未获得有效地址的图像块(即 HIT(j)=0)，则在散列地址空间中查找首个未被占用的位置 α，分配该地址作为 HAT(j) 的对应值(即 HAT(j)=α)并标记 OT(α)=1。

根据所有图像块序列号 j，将其所映射的散列地址 HAT(j) 反向记录到 IHAT 中，即设置 IHAT(HAT(j))=j。

步骤 3 生成认证数据 $D_A$

对于每个图像块 $B_j$，将其在 HIT 中的值(即 HIT(j)=k 表示其所选用的哈希函数编号)及所对应的奇偶校验位共同组成认证数据 $D_A$，用于提高篡改检测的鲁棒性与准确性。

步骤 4 生成恢复数据 $D_R$

对原始图像 I 应用 SPIHT 编码算法并控制截断速率为 1 bpp，以获取压缩比高、分辨率保留性强的码流。针对每个 2×2 大小的图像块 $B_j$，从 SPIHT 输出流中分配 4bits 编码数据作为其自身恢复信息并记为 $D_{RS}(j)$。根据 IHAT 关系，将图像块 $B_j$ 的逆映射块 $B_{IHAT(j)}$ 的 SPIHT 输出流中 4bits 部分作为其逆恢复信息并记为 $D_{RM}(j)$。最后，将二者组合成完整的恢复数据 $D_R(j)=D_{RS}(j)\parallel D_{RM}(j)$。

### 3.2 基于层级扩展的脆弱水印算法

目前，已有大量研究将空间域的水印嵌入方法[24-28]应用于图像篡改检测与自恢复领域。然而，这类方法在实际应用中仍面临两个核心挑战：一方面，为了提高篡改检测的精度，通常需要嵌入大量认证信息，这导致嵌入容量受到限制；另一方面，在提升嵌入容量的同时往往会导致图像失真，影响水印图像的视觉质量。如何在确保高检测精确度的前提下兼顾嵌入容量与图像质量，成为当前研究中的关键问题。在本方案中是由融合完美哈希与 SPIHT 的图像块级匹配算法用于生成认证数据 $D_A$ 与恢复数据 $D_R$，并将其作为水印信息嵌入至原始图像中，根据上述算法流程可知每个 2×2 图像块 $B_j$ 至少需要嵌入 12bits 的水印信息，因此其对应的嵌入率应不低于 3bpp(bits per pixel)。针对这些问题，本文提出了一种基于层级扩展的脆弱水印算法，用于实现所提方案的水印嵌入过程，以在满足高嵌入容量的同时保持较好的图像质量。

在水印嵌入过程中，该算法首先根据式(5)构建一个大小为 256×256 的层级扩展矩阵 $\boldsymbol{M}$，如图 3 所示，其中矩阵的所有元素均为 0 到 8 的九进制数字。此外，该矩阵的构建公式依赖于嵌入密钥 $K$，该密钥可通过 QKD 技术实现安全传输，从而进一步增强算法的安全性。

$$\boldsymbol{M}(p_i, p_{i+1}) = \left(\left\lfloor \frac{p_{i+1} \bmod 27}{3} \right\rfloor + (p_i \bmod 9) + 3\cdot(p_{i+1} \bmod 3) - 1\right) \bmod 9 \tag{5}$$

矩阵 $\boldsymbol{M}$ 具有层级扩展结构，可按 3×3 大小将其划分成子单元 $U(q_x, q_y)$，每个子单元包含 0 到 8 的不重复九进制数。将子单元 $U(q_x, q_y)$ 在矩阵 $\boldsymbol{M}$ 上所对应位置的中心元素作为该子单元的代表值，由式(6)计算。同样地，每个像素对 $(p_i, p_{i+1})$ 都属于一个 3×3 子单元 $U(q_x, q_y)$，其中 $q_x=\lfloor p_i/3 \rfloor$ 且 $q_y=\lfloor p_{i+1}/3 \rfloor$。

$$U(q_x, q_y) = M(3\cdot q_x + 1, 3\cdot q_y + 1) \tag{6}$$

所有 3×3 子单元 $U(q_x, q_y)$ 的中心元素可以重新组成一个大小为 85×85 二维空间的子矩阵 $\boldsymbol{M}'$，如图 4 所示，在该矩阵中以任意元素为中心所组成的新 3×3 单元 $U'$ 可由式(7)表示，其均为 0 到 8 的不重复九进制数字。

$$U'(q_x, q_y) = \{U(q_x + v_1, q_y + v_2) \mid v_1, v_2 \in \{-1, 0, 1\}\} \tag{7}$$

水印信息的嵌入过程如算法 2 所示。

**算法 2** 水印信息的嵌入算法

输入：h×w 的原始图像 I，二进制信息流 S，嵌入密钥 K

输出：水印图像 $I_D$

步骤 1 预处理与进制转换

将原始图像 I 按顺序扫描并划分成一维不重叠的原始像素

对序列$(p_i,p_{i+1})$，其中$p_i$和$p_{i+1}$代表两个相邻原始像素的灰度值。根据嵌入密钥K构建式(5)对应的256×256大小的层级扩展矩阵**M**。将二进制信息流S转换成九进制信息$S_n=(s_1,s_2,\cdots,s_l)_9$，其中$l=h\times w\times\log_2 9$表示信息长度。

步骤2　确定原始元素位置

设置i=1，将像素对$(p_i,p_{i+1})$映射到层级扩展矩阵**M**的对应位置$M(p_i,p_{i+1})$，确定其所属的子单元$U(q_x,q_y)$，即$U(\lfloor p_i/3\rfloor,\lfloor p_{i+1}/3\rfloor)$，根据式(6)可知该子单元的代表元素值为$M(3\cdot\lfloor p_i/3\rfloor+1,3\cdot\lfloor p_{i+1}/3\rfloor+1)$。

步骤3　嵌入信息

按序检索每两位九进制信息$(s_i,s_{i+1})_9$，将其嵌入到原始图像I对应的像素对$(p_i,p_{i+1})$中。首先，嵌入第一位九进制信息$s_i$。若$s_i$等于$M(p_i,p_{i+1})$所属子单元$U(q_x,q_y)$的代表值，则令$M(\hat{p_i},\widehat{p_{i+1}})=M(3\cdot\lfloor p_i/3\rfloor+1,3\cdot\lfloor p_{i+1}/3\rfloor+1)=s_i$；否则，在以该子单元$U(q_x,q_y)$元素为中心的二维空间子矩阵**M′**中，根据式(5)找到其中一个元素能满足$s_i=U(q_x',q_y')$，并确定替代位置为$M(\hat{p_i},\widehat{p_{i+1}})=M(3\cdot q_x'+1,3\cdot q_y'+1)=s_i$。接着，嵌入第二位九进制信息$s_{i+1}$。若$s_{i+1}=M(\hat{p_i},\hat{p_{i+1}})$，则保持不变，即最终修改位置$M(p_i',p_{i+1}')=M(\hat{p_i},\widehat{p_{i+1}})$；否则在$M(\hat{p_i},\widehat{p_{i+1}})$所属的子单元中找到一个元素能满足$s_{i+1}=M(p_i',p_{i+1}')$并确立最终修改位置。

步骤4　循环

令i=i+2，重复步骤2－步骤3，直到所有秘密信息嵌入完毕，则输出水印图像$I_D$。

接收方在收到水印图像$I_D$后，可以根据共享的嵌入密钥$K$通过算法3的步骤精确提取嵌入的二进制信息流$S$。水印信息的提取过程如算法3所示。

**算法3**　水印信息提取算法

输入：h×w的水印图像$I_D$，嵌入密钥K

输出：二进制信息流S

步骤1　预处理

将水印图像$I_D$按顺序扫描并划分成一维不重叠的像素对序列$(p_i',p_{i+1}')$，其中$p_i'$和$p_{i+1}'$代表两个相邻水印像素的灰度值，并初始化i=1。根据共享的嵌入密钥K构建式(5)所对应的256×256大小的层级扩展矩阵**M**。

步骤2　提取信息

将像素对$(p_i',p_{i+1}')$映射到层级扩展矩阵**M**的对应位置**M**$(p_i',p_{i+1}')$，确定其所属的子单元$U(q_x,q_y)$即$U(\lfloor p_i'/3\rfloor,\lfloor p_{i+1}'/3\rfloor)$。则嵌入的两位九进制数字满足如下计算式：

$$s_i=M(3\cdot\lfloor p_i'/3\rfloor+1,3\cdot\lfloor p_{i+1}'/3\rfloor+1),\ s_{i+1}=M(p_i',p_{i+1}') \tag{8}$$

步骤3　循环并输出结果

令i=i+2，重复步骤2直到水印图像$I_D$中的所有像素对均被遍历，则秘密信息提取完毕。最后，将所有提取的九进制信息通过进制转换还原为原始二进制信息流S。

$P_{i+1}$

| $p_i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | … | … | 254 | 255 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 8 | 2 | 5 | 0 | 3 | 6 | 1 | 4 | 7 | 2 | 5 | 8 | 3 | 6 | 0 | 4 | 7 | 1 | 5 | 8 | 2 | 6 | 0 | 3 | 7 | 1 | 4 | 8 | 2 | … | … | 8 | 3 |
| 1 | 0 | **3** | 6 | 1 | **4** | 7 | 2 | **5** | 8 | 3 | **6** | 0 | 4 | **7** | 1 | 5 | **8** | 2 | 6 | **0** | 3 | 7 | **1** | 4 | 8 | **2** | 5 | 0 | 3 | … | … | 0 | 4 |
| 2 | 1 | 4 | 7 | 2 | 5 | 8 | 3 | 6 | 0 | 4 | 7 | 1 | 5 | 8 | 2 | 6 | 0 | 3 | 7 | 1 | 4 | 8 | 2 | 5 | 0 | 3 | 6 | 1 | 4 | … | … | 1 | 5 |
| 3 | 2 | 5 | 8 | 3 | 6 | 0 | 4 | 7 | 1 | 5 | 8 | 2 | 6 | 0 | 3 | 7 | 1 | 4 | 8 | 2 | 5 | 0 | 3 | 6 | 1 | 4 | 7 | 2 | 5 | … | … | 2 | 6 |
| 4 | 3 | **6** | 0 | 4 | **7** | 1 | 5 | **8** | 2 | 6 | **0** | 3 | 7 | **1** | 4 | 8 | **2** | 5 | 0 | **3** | 6 | 1 | **4** | 7 | 2 | **5** | 8 | 3 | 6 | … | … | 3 | 7 |
| 5 | 4 | 7 | 1 | 5 | 8 | 2 | 6 | 0 | 3 | 7 | 1 | 4 | 8 | 2 | 5 | 0 | 3 | 6 | 1 | 4 | 7 | 2 | 5 | 8 | 3 | 6 | 0 | 4 | 7 | … | … | 4 | 8 |
| 6 | 5 | 8 | 2 | 6 | 0 | 3 | 7 | 1 | 4 | 8 | 2 | 5 | 0 | 3 | 6 | 1 | 4 | 7 | 2 | 5 | 8 | 3 | 6 | 0 | 4 | 7 | 1 | 5 | 8 | … | … | 5 | 0 |
| 7 | 6 | **0** | 3 | 7 | **1** | 4 | 8 | **2** | 5 | 0 | **3** | 6 | 1 | **4** | 7 | 2 | **5** | 8 | 3 | **6** | 0 | 4 | **7** | 1 | 5 | **8** | 2 | 6 | 0 | … | … | 6 | 1 |
| 8 | 7 | 1 | 4 | 8 | 2 | 5 | 0 | 3 | 6 | 1 | 4 | 7 | 2 | 5 | 8 | 3 | 6 | 0 | 4 | 7 | 1 | 5 | 8 | 2 | 6 | 0 | 3 | 7 | 1 | … | … | 7 | 2 |
| 9 | 8 | 2 | 5 | 0 | 3 | 6 | 1 | 4 | 7 | 2 | 5 | 8 | 3 | 6 | 0 | 4 | 7 | 1 | 5 | 8 | 2 | 6 | 0 | 3 | 7 | 1 | 4 | 8 | 2 | … | … | 8 | 3 |
| 10 | 0 | 3 | 6 | 1 | 4 | 7 | 2 | 5 | 8 | 3 | 6 | 0 | 4 | 7 | 1 | 5 | 8 | 2 | 6 | 0 | 3 | 7 | 1 | 4 | 8 | 2 | 5 | 0 | 3 | … | … | 0 | 4 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 254 | 1 | 4 | 7 | 2 | 5 | 8 | 3 | 6 | 0 | 4 | 7 | 1 | 5 | 8 | 2 | 6 | 0 | 3 | 7 | 1 | 4 | 8 | 2 | 5 | 0 | 3 | 6 | 1 | 4 | … | … | 1 | 5 |
| 255 | 2 | 5 | 8 | 3 | 6 | 0 | 4 | 7 | 1 | 5 | 8 | 2 | 6 | 0 | 3 | 7 | 1 | 4 | 8 | 2 | 5 | 0 | 3 | 6 | 1 | 4 | 7 | 2 | 5 | … | … | 2 | 6 |

图3　层级扩展矩阵 ***M***

Fig. 3　Hierarchical extended matrix ***M***

$q_y$

| $q_x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | … | … | 82 | 83 | 255 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | … | … | 7 | 8 | 0 |
| 1 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | … | … | 1 | 2 | 3 |
| 2 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | … | … | 4 | 5 | 6 |
| 3 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | … | … | 7 | 8 | 0 |
| 4 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | … | … | 1 | 2 | 3 |
| 5 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | … | … | 4 | 5 | 6 |
| 6 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | … | … | 7 | 8 | 0 |
| 7 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | … | … | 1 | 2 | 3 |
| 8 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | … | … | 4 | 5 | 6 |
| 9 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | … | … | 7 | 8 | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 82 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | … | … | 1 | 2 | 3 |
| 83 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | … | … | 4 | 5 | 6 |
| 84 | 3 | 4 | 5 | 6 | 7 | 8 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | … | … | 7 | 8 | 0 |

图4　子矩阵 ***M′***

Fig. 4　Sub-matrix ***M′***

### 3.3　篡改检测与自恢复算法

为实现对图像内容在公开信道传输过程中可能遭受的恶意篡改进行精确识别与自恢复，本文提出以已嵌入认证数据与恢复数据的脆弱水印图像为基础，通过认证数据检测定位篡改区域，并结合逆映射恢复位对被破坏的图像区域进行恢复。由于图像中篡改区域内嵌入的水印信息会随着内容的破坏而丢失，本算法设计了包含图像块自身恢复信息与其逆映射块恢复信息的双向恢复机制，有效增强了图像恢复的鲁棒性。该算法适用于电力图像传输等安全敏感场景，具有良好的检测精度与自恢复性能。算法的流程如算法4所示。

**算法4**　篡改检测与自恢复算法

输入：待认证图像I′，量子随机数种子r，共享密钥K

输出：篡改检测图像$I_t$，自恢复图像$I_r$

步骤1　初始化

将待认证图像I′划分为不重叠的2×2大小图像块$B_j$，其中总图像块数为h×w/4。创建长度为h×w/4的篡改检测表(Tamper Detection Table，TDT)和重构定位表(Reconstitution Location Table，RLT)将表内的值全部初始化为0。TDT用于记录图像块$B_j$是否被篡改，RLT用于存储每个图像块的重构数据。

步骤 2 预处理

利用共享的嵌入密钥 K 构建式(5)所对应的 256×256 大小的层级扩展矩阵 **M**，并根据 3.2 节的水印信息提取算法，从待认证图像 I′中精确提取认证数据 $D_A'$ 与恢复数据 $D_R'$。基于量子随机数种子 r 构造与发送端一致的完美哈希，根据 3.1 节所述构建哈希映射关系，计算每个图像块 $B_j$ 对应的 HIT 特征位，并将其与所对应的奇偶校验位共同组成参考的认证数据 $D_A$，用于检测并定位篡改区域。

步骤 3 篡改检测

将每个图像块 $B_j$ 所提取的对应认证数据 $D_A'$ 与计算所得参考认证数据 $D_A$ 进行对比，若两者一致，则表示该区域未被篡改，保持 TDT(j)的值不变；否则设置 TDT(j)=1，表示该区域已被篡改。

步骤 4 可视化

令 j=j+1，重复步骤 3 直到 j=h×w/4 结束，从而生成待认证图像 I′对应的 TDT，根据 TDT 对篡改检测结果进行可视化，篡改检测图像 $I_t$ 由式(9)计算所得，为了进一步降低误检率，引入形态学处理操作，包括图像腐蚀以去除孤立噪声、闭运算填充封闭孔洞，以及边缘扩张以消除残留空洞等。经优化后生成最终的篡改检测图像 $I_t$。

$$I_t\left(2\cdot\left\lfloor\frac{u-1}{w/2}\right\rfloor+v,2\cdot((u-1)\bmod\left(\frac{w}{2}\right))+v\right)=TDT(j),\quad v\in\{1,2\} \tag{9}$$

步骤 5 恢复图像

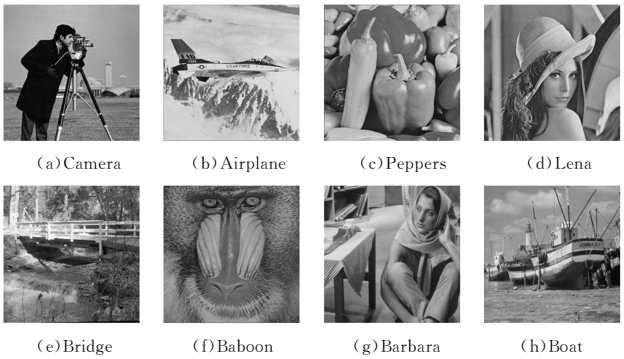
对于每个图像块 $B_j$ 所提取的恢复数据 $D_R'$，将其划分为图像块 $B_j$ 自身的恢复位 $D_{RS}$ 和及其逆映射图像块 $B_{IHAT(j)}$ 的恢复位 $D_{RM}$。再根据步骤 4 计算所得 TDT 中第 j 个图像块对应的检测结果，动态决定该图像块所采用的恢复信息。若 $B_j$ 被判定未篡改，则直接采用其嵌入的自恢复位 $D_{RS}$ 作为当前图像块的重构数据；否则使用其映射图像块 $B_{HAT(j)}$ 中所嵌入的逆映射恢复位 $D_{RM}$ 作为重构数据。其计算式如下：

$$RLT(j)=\begin{cases}D_{RS}(j), & \text{if } TDT(j)=0\\ D_{RM}(HAT(j)), & \text{if } TDT(j)=1\end{cases} \tag{10}$$

最后，将 RLT 中的数据转换为比特流形式的重构数据 B，并通过 SPIHT 解码算法进行图像重构。为进一步提升视觉质量，本算法将解码图像作为修复源，仅替换待认证图像 I′中检测为篡改的区域，以获得结构一致、视觉较好的恢复图像 $I_r$。

## 4 实验结果及分析

衡量基于脆弱水印的图像篡改检测与自恢复方法性能的常用指标包括水印嵌入容量、图像视觉质量、篡改检测精度、自恢复能力等[29]。为验证本文所提出方案的有效性，实验选取 UCS-SIPI 图像数据库中分辨率为 512×512 的标准灰度图像作为测试样本，所有仿真实验均在 MATLAB 软件中完成。



(a)Camera (b)Airplane (c)Peppers (d)Lena

(e)Bridge (f)Baboon (g)Barbara (h)Boat

图 5 标准灰度图像

Fig. 5 Standard gray image

### 4.1 水印嵌入容量与视觉质量

峰值信噪比(Peak Signal-to-Noise Ratio，PSNR)是衡量水印图像质量的核心指标，用于评估原始图像与水印图像之间的相似程度，单位为 dB。

$$PSNR=10\times\log_{10}\frac{255^2}{MSE} \tag{11}$$

其中，MSE 表示原始图像与水印图像之间的均方误差，计算式为：

$$MSE=\frac{1}{H\times W}\times\sum_{i=1}^{W}\sum_{j=1}^{H}(I(i,j)-I'(i,j))^2 \tag{12}$$

其中，$I(i,j)$ 和 $I'(i,j)$ 分别为原始图像与水印图像在像素位置 $(i,j)$ 的灰度值，$H$ 和 $W$ 分别为图像的高度和宽度。若两幅图像越相似，MSE 越小，PSNR 越大，则说明水印嵌入后对视觉质量的影响较小；反之，则表明失真程度较高。除 PSNR 外，结构相似性(Structural Similarity Index，SSIM)和质量指数(Quality Index，QI)也是评价水印图像质量的重要指标，分别定义为：

$$SSIM=\frac{(2\mu_i\mu_j+c_1)(2\sigma_{ij}+c_2)}{(\mu_i^2+\mu_j^2+c_1)(\sigma_i^2+\sigma_j^2+c_2)},0\leqslant SSIM\leqslant 1 \tag{13}$$

$$QI=\frac{4\sigma_{ij}\mu_i\mu_j}{(\sigma_i^2+\sigma_j^2)(\mu_i^2+\mu_j^2)},0\leqslant QI\leqslant 1 \tag{14}$$

其中，$\mu_i$ 和 $\mu_j$ 为原始图像与水印图像的均值，$\sigma_i^2$ 和 $\sigma_j^2$ 为其方差，$\sigma_{ij}$ 表示协方差，$c_1=(t_1R)^2$，$c_2=(t_2R)^2$，其中 $t_1=0.01$，$t_2=0.03$，$R$ 为图像灰度值的动态范围。SSIM 和 QI 越接近 1，表示水印嵌入后对图像结构和感知质量的影响越小。

信息嵌入率(Embedding Rate，ER)用于衡量水印的有效载荷，定义为单位像素所携带的比特数，单位为 bpp(bits per pixel)。

$$ER=\frac{\|S\|}{H\times W} \tag{15}$$

其中，$\|S\|$ 表示嵌入的比特总数，也可称之为嵌入容量(Embedding Capacity，EC)；$H$ 和 $W$ 为图像尺寸。较高的 ER 意味着更大的嵌入容量，但也会影响图像质量，因此性能优秀的数字水印算法应在保证较高的信息嵌入率 ER 和嵌入容量 EC 的同时，尽可能保持较好的 PSNR，MSE，SSIM 和 QI 等质量指标。表 1 列出了本文所提的基于层次扩展的脆弱水印算法在嵌入容量与视觉质量方面的实验结果。

表 1 脆弱水印图像的实验结果

Table 1 Experimental results of fragile watermark images

| Images | *PSNR*/dB | *MSE* | *SSIM* | *QI* | *ER*/bpp | *EC*/bits |
|---|---|---|---|---|---|---|
| Camera | 39.61 | 7.10 | 0.9925 | 0.9988 | 3.17 | 803976 |
| Airplane | 39.64 | 7.06 | 0.9870 | 0.9984 | 3.17 | 803976 |
| Peppers | 39.62 | 7.09 | 0.9926 | 0.9985 | 3.17 | 803976 |
| Lena | 39.65 | 7.04 | 0.9890 | 0.9985 | 3.17 | 803976 |
| Bridge | 39.65 | 7.05 | 0.9908 | 0.9984 | 3.17 | 803976 |
| Baboon | 39.63 | 7.08 | 0.9887 | 0.9988 | 3.17 | 803976 |
| Barbara | 39.64 | 7.05 | 0.9948 | 0.9973 | 3.17 | 803976 |
| Boat | 39.63 | 7.08 | 0.9900 | 0.9983 | 3.17 | 803976 |
| average | 39.63 | 7.07 | 0.9907 | 0.9984 | 3.17 | 803976 |

表 2 列出了本文所提出的基于层次扩展的脆弱水印算法与现有图像篡改检测方案中所采用的水印算法[15-16]的性能。实验结果显示，本文算法在嵌入载荷方面相较于文献[15-16]分别提高了 0.17 bpp 和 1.67 bpp，且具有较高的图像视觉质量。

表 2　水印图像质量与嵌入载荷的对比

Table 2　Comparison of watermark image quality and embedding load

| Images | Sreenivas et al. scheme[15] | | Prasad et al. scheme[16] | | Proposed scheme | |
|---|---|---|---|---|---|---|
| | *PSNR*/dB | *ER*/bpp | *PSNR*/dB | *ER*/bpp | *PSNR*/dB | *ER*/bpp |
| Camera | 37.92 | 3 | 42.29 | 1.5 | 39.61 | 3.17 |
| Airplane | 38.06 | 3 | 42.95 | 1.5 | 39.64 | 3.17 |
| Peppers | 37.93 | 3 | 42.21 | 1.5 | 39.62 | 3.17 |
| Lena | 37.98 | 3 | 42.01 | 1.5 | 39.65 | 3.17 |
| Bridge | 37.91 | 3 | 42.23 | 1.5 | 39.65 | 3.17 |
| Baboon | 37.89 | 3 | 41.82 | 1.5 | 39.63 | 3.17 |
| Barbara | 37.93 | 3 | 42.29 | 1.5 | 39.64 | 3.17 |
| Boat | 37.98 | 3 | 41.11 | 1.5 | 39.63 | 3.17 |
| average | 37.85 | 3 | 42.11 | 1.5 | 39.63 | 3.17 |

### 4.2　篡改检测精度与自恢复性能

篡改检测性能通常通过以下 4 个主要指标进行衡量：精确度（Precision）、召回率（Recall）、误检率（False Detection Rate，FDR）和误警率（False Alarm Rate，FAR）。其中，Precision 和 Recall 的计算式分别如下：

$$Precision=\frac{TP}{TP+FP} \tag{16}$$

$$Recall=\frac{TP}{TP+FN} \tag{17}$$

其中，$TP$（True Positive）为被正确检测为篡改区域的像素数量，$FP$（False Positive）为被错误检测为篡改区域的像素数量，$FN$（False Negative）为未能检测出的篡改区域像素数量，$TN$（True Negative）为被正确检测为未篡改区域的像素数量。FDR 和 FAR 的计算式分别如下：

$$FDR=\frac{FN}{TP+FN} \tag{18}$$

$$FAR=\frac{FP}{FP+TN} \tag{19}$$

显然，当 Precision 和 Recall 越接近 100%，而 FDR 和 FAR 越接近 0 时，表明该图像认证方法在篡改检测方面的性能越优异。

图 6 给出了本文所提出的图像篡改检测与自恢复算法在多种典型场景下的仿真结果。实验选取了具有不同纹理特征的 4 幅标准图像（Camera，Airplane，Lena 和 Barbara），在其平滑区域、粗糙纹理区以及显著边缘区进行了有针对性的篡改处理，以全面评估本文算法在不同复杂度图像下的适应性和鲁棒性。图 6 第 1 行为嵌入认证数据与恢复数据后的脆弱水印图像。与图 5 中的原始图像相比，尽管嵌入了额外的数据载荷，水印图像仍保持了良好的视觉质量，表明所提出算法在嵌入阶段具有较高的隐蔽性。第 2 行展示了对应的篡改图像。第 3 行和第 4 行分别展示了本文的篡改检测算法定位到的篡改区域结果及其恢复图像。实验结果表明，无论篡改区域呈规则还是不规则形态，本文算法均能够准确识别篡改位置并实现有效恢复。尤其是在图 6(c-2)中，尽管 Lena 图像经过高达 50%的大面积篡改，导致原始内容几乎完全不可辨，但本文算法仍能够实现较为完整的图像内容恢复，保障了恢复图像的结构完整性和视觉质量。

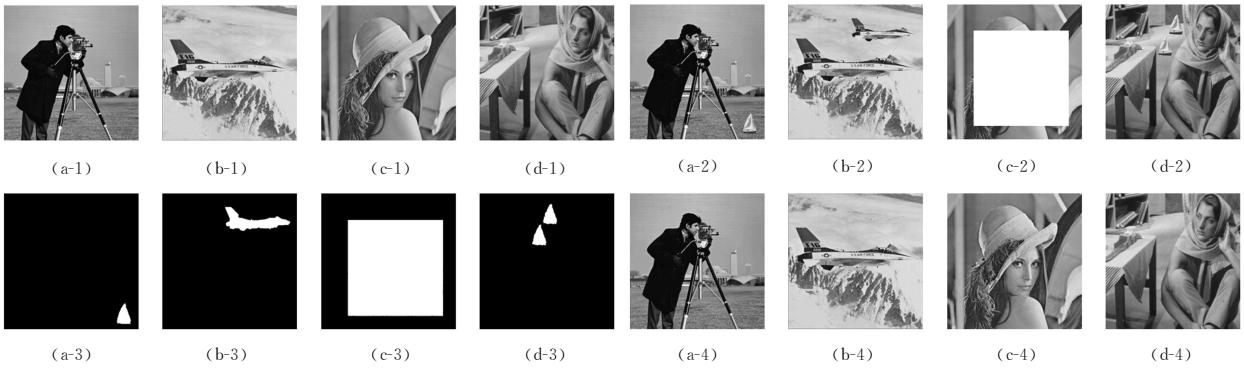



图 6　篡改检测与自恢复的仿真结果

Fig. 6　Simulation results of tamper detection and self-recovery

此外，表 3 对比了本文方法与现有的图像认证方案[30-31]在篡改检测性能方面的表现。实验结果表明，本文方法在各项性能指标上均优于上述方法[30-31]，进一步验证了该方案在实际应用中的有效性与可行性。

表 3　图像篡改检测算法的量化对比

Table 3　Quantitative comparison of image tampering detection algorithms

（%）

| Schemes | *Precision* | *Recall* | *FDR* | *FAR* |
|---|---|---|---|---|
| Peng 等[30] | 98.51 | 97.38 | 2.62 | 0.02 |
| Yin 等[31] | 95.27 | 90.79 | 9.21 | 0.06 |
| Proposed | 99.01 | 98.31 | 1.69 | 0.01 |

**结束语**　本文针对电力系统远程监控与无人值守巡检等典型应用场景中图像数据易受篡改与伪造的问题，提出了一种基于量子安全和脆弱水印的图像篡改与自恢复算法。该算法通过构造基于量子随机数的完美哈希认证模型，结合图像块级匹配策略与 SPIHT 编码，有效生成认证与恢复数据，并以脆弱水印形式嵌入到原始通信图像中，从而在保证信息隐蔽性的同时，显著提升了篡改检测的准确性与被篡改区域的自恢复能力。实验结果表明，所提方案在篡改检测精度、自恢复效果、安全性、嵌入容量和图像质量等关键性能指标上均优于现有同类方法，验证了其在电力图像传输完整性保护与可信认证中的可行性与优越性。

综上所述，随着电力系统向智能化、信息化不断发展，针对网络通信中图像数据面临的篡改与伪造风险，本文提出的方案提供了一种可行且高效的安全保障机制。未来，可进一步探索该方法在更多实际场景下的应用，并结合新兴的量子通信技术与人工智能算法，不断提升电力系统网络的安全防护水平与应急处置能力，以解决在实际应用中的量子密钥分发效率、随机数生成速率以及系统实时性等方面的挑战问题。具体可包括：1）结合新型量子通信网络，优化量子密钥分发的可扩展性与抗干扰能力；2）探索面向复杂电力通信环境下的

大规模多图像并行认证与快速自恢复机制，以满足电力物联网对高效性与鲁棒性的更高要求；3)将人工智能技术引入篡改检测与恢复流程，提升对多样化篡改手段的适应性与检测准确度。

## 参 考 文 献


[1] MADHUSHREE B, BASANTH K H B, CHENNAMMA H R. An exhaustive review of authentication, tamper detection with localization and recovery techniques for medical images[J]. Multimedia Tools and Applications, 2024, 83(13): 39779-39821.

[2] ALORAINI M, SHARIFZADEH M, SCHONFELD D. Sequential and patch analyses for object removal video forgery detection and localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(3): 917-930.

[3] SHAIK A S, KARSH R K, ISLAM M, et al. A Secure and Robust Autoencoder-Based Perceptual Image Hashing for Image Authentication[J]. Wireless Communications and Mobile Computing, 2022, 16(4): 56-58.

[4] BAPPY J H, SIMONS C, NATARAJ L, et al. Hybrid LSTM and encoder-decoder architecture for detection of image forgeries [J]. IEEE Transactions on Image Processing, 2019, 28(7): 3286-3300.

[5] BHALERAO S, ANSARI I A, KUMAR A. A secure image watermarking for tamper detection and localization[J]. Journal of Ambient Intelligence and Humanized Computing, 2021, 12(1): 1057-1068.

[6] SINHAL R, ANSARI I A. Multipurpose image watermarking: ownership check, tamper detection and self-recovery[J]. Circuits, Systems, and Signal Processing, 2022, 41(6): 3199-3221.

[7] VAIDYA S P, KANDALA R N, CHANDRA M P, et al. A robust fragile watermarking approach for image tampering detection and restoration utilizing hybrid transforms[J]. Scientific Reports, 2025, 15(1): 1-24.

[8] AHMADI S B B, ZHANG G, RABBANI M, et al. An intelligent and blind dual color image watermarking for authentication and copyright protection[J]. Applied Intelligence, 2021, 51(3): 1701-1732.

[9] VAN S R G, TIRKEL A Z, OSBORNE C F. A digital watermark[C]// Proceedings of 1st International Conference on Image Processing. 1994: 86-90.

[10] LIN P L, HSIEH C K, HUANG P W. A hierarchical digital watermarking method for image tamper detection and recovery[J]. Pattern Recognition, 2005, 38(12): 2519-2529.

[11] LEE T Y, LIN S D. Dual watermark for image tamper detection and recovery[J]. Pattern Recognition, 2008, 41(11): 3497-3506.

[12] QIN C, JI P, ZHANG X, et al. Fragile image watermarking with pixel-wise recovery based on overlapping embedding strategy [J]. Signal Processing, 2017, 138: 280-293.

[13] SARRESHTEDARI S, AKHAEE M A. A source-channel coding approach to digital image protection and self-recovery[J]. IEEE Transactions on Image Processing, 2015, 24(7): 2266-2277.

[14] BOLOURIAN H B, TAHERINIA A H, MOHAJERZADEH A H. TRLG: Fragile blind quad watermarking for image tamper detection and recovery by providing compact digests with optimized quality using LWT and GA[J]. Information Sciences, 2019, 486: 204-230.

[15] SREENIVAS K, KAMAKSHIPRASAD V. Improved image tamper localisation using chaotic maps and self-recovery[J]. Journal of Visual Communication and Image Representation, 2017, 49: 164-176.

[16] PRASAD S, PAL A K. A tamper detection suitable fragile watermarking scheme based on novel payload embedding strategy [J]. Multimedia Tools and Applications, 2020, 79(3): 1673-1705.

[17] LIU T, YUAN X. Adaptive feature calculation and diagonal mapping for successive recovery of tampered regions[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(7): 2617-2630.

[18] CAO Y, ZHAO Y, WANG Q, et al. The evolution of quantum key distribution networks: On the road to the qinternet[J]. IEEE Communications Surveys & Tutorials, 2022, 24(2): 839-894.

[19] MANNALATHA V, MISHRA S, PATHAK A. A comprehensive review of quantum random number generators: Concepts, classification and the origin of randomness[J]. Quantum Information Processing, 2023, 22(12): 439-445.

[20] SPRUGNOLI R. Perfect hashing functions: a single probe retrieving method for static sets[J]. Communications of the ACM, 1977, 20(11): 841-850.

[21] DU M W, HSIEH T M, JEA K F, et al. The study of a new perfect hash scheme[J]. IEEE Transactions on Software Engineering, 1983, 9(3): 305-313.

[22] SHAPIRO J. Embedded image coding using zerotrees of wavelet coefficients[J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3445-3462.

[23] SAID A, PEARLMAN W A. A new, fast, and efficient image codec based on set partitioning in hierarchical trees[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1996, 6(3): 243-250.

[24] BENDER W, GRUHL D, MORIMOTO N, et al. Techniques for data hiding[J]. IBM Systems Journal, 1996, 35(34): 313-336.

[25] LIU Y X, YANG C N, SUN Q D, et al. Enhanced embedding capacity for the SMSD-based data-hiding method[J]. Signal Processing: Image Communication, 2019, 78: 216-222.

[26] HE W, CAI Z. Reversible data hiding based on dual pairwise prediction-error expansion[J]. IEEE Transactions on Image Processing, 2021, 30: 5045-5055.

[27] SHEN S Y, HUANG L H. A data hiding scheme using pixel value differencing and improving exploiting modification directions[J]. Computers & Security, 2015, 48: 131-141.

[28] LIU Y J, CHANG C C, HUANG P C. Extended exploiting-modification-direction data hiding with high capacity[C]// Proceedings of the International Conference on Video and Image Processing. ACM, 2017: 151-155.

[29] HAOUZIA A, NOUMEIR R. Methods for image authentication: a survey[J]. Multimedia Tools and Applications, 2008, 39(1): 1-46.

[30] PENG Y, NIU X, FU L, et al. Image authentication scheme based on reversible fragile watermarking with two images[J]. Journal of Information Security and Applications, 2018, 40: 236-246.

[31] YIN Z, NIU X, ZHOU Z, et al. Improved reversible image authentication scheme[J]. Cognitive Computation, 2016, 8(5): 890-899.


**CHEN Hongxiang**, born in 1995, master, senior engineer. His main research interests include image security, quantum cryptography and hydropower station cybersecurity.