中国科学技术大学
University of Science and Technology of China

博士学位论文答辩

3D隐写模型与方法研究

答辩人：周 航
导师：俞能海教授、张卫明教授
学科专业：网络空间安全

2020年6月

0. 报告目录

中国科学技术大学
University of Science and Technology of China

目录

1 研究背景

信息时代中，隐私数据的盗窃造成大量损失。

2020年1月，日本的三菱电机确认在2019年6月发生数据泄露事件，导致200MB敏感数据被盗。

MITSUBISHI ELECTRIC
Changes for the Better

2020年1月，Dixon Carphones公司数据保护措施不足，导致560万支付卡信息和1400万用户信息被窃取。

Dixons Carphone

2020年3月，黑暗阅读报道5.38亿微博用户的个人信息已经在暗网上被以250美元的价格出售，数据包括用户名、位置、电话号码等。

新浪微博 weibo.com

2020年3月，Open Exchange Rates宣布黑客从非法访问了部分客户的个人信息和哈希密码。

open exchange rates

1 研究背景

如何保证隐私信息的安全性?

1 研究背景

如何保证隐私信息的安全性?

- 加密：追求内容安全，与真随机不可区分
- 隐写：追求行为安全，与载体不可区分，具备伪装性

1 研究背景

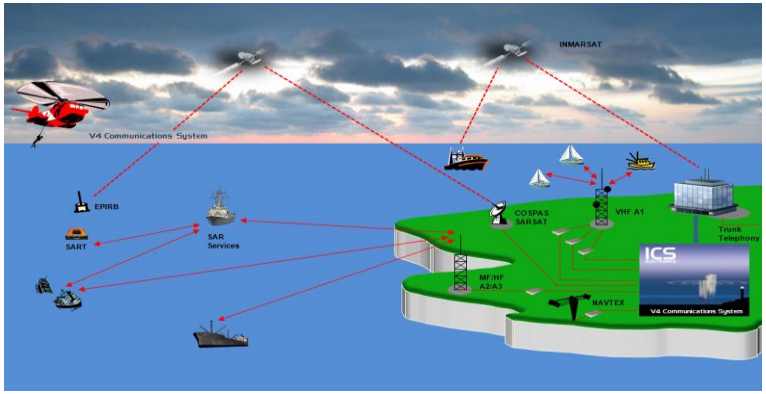
如何保证隐私信息的安全性?

- 加密：追求内容安全，与真随机不可区分
- 隐写：追求行为安全，与载体不可区分，具备伪装性

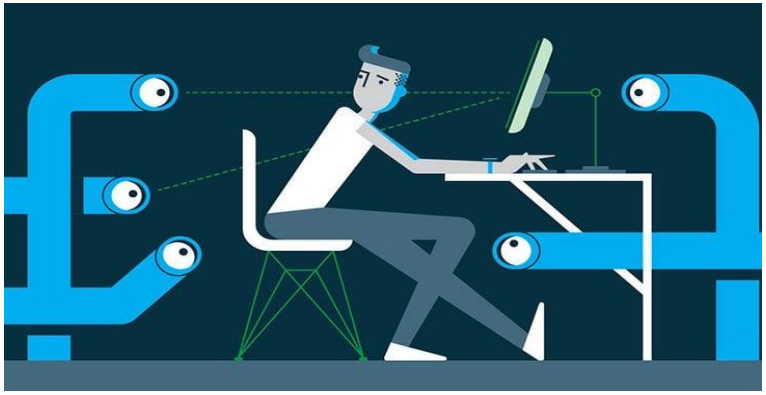
1 研究背景

数字隐写应用场景

- 隐蔽通信

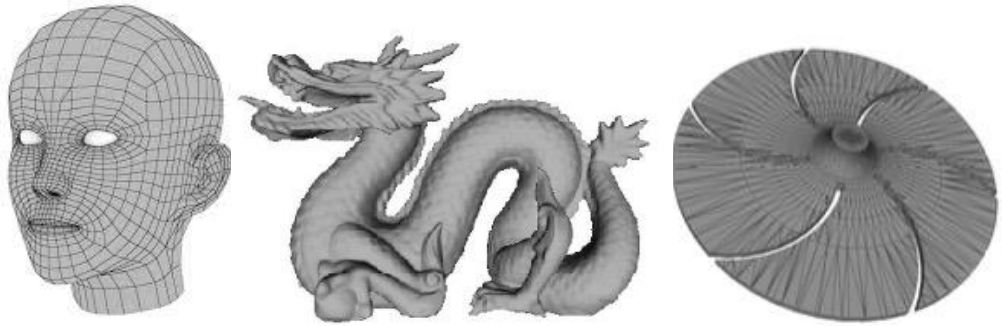


- 个人隐私保护

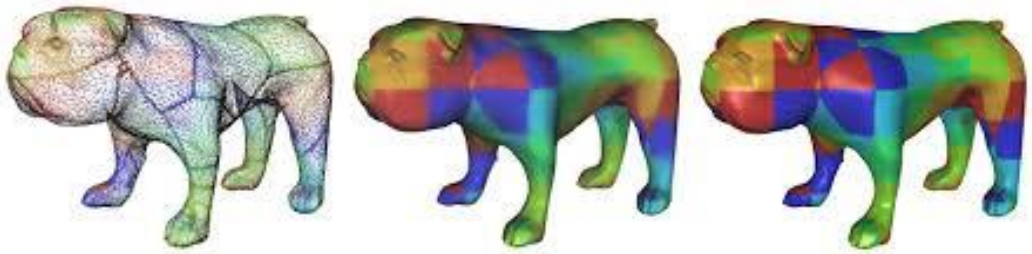


- 军事网络攻防

1 研究背景



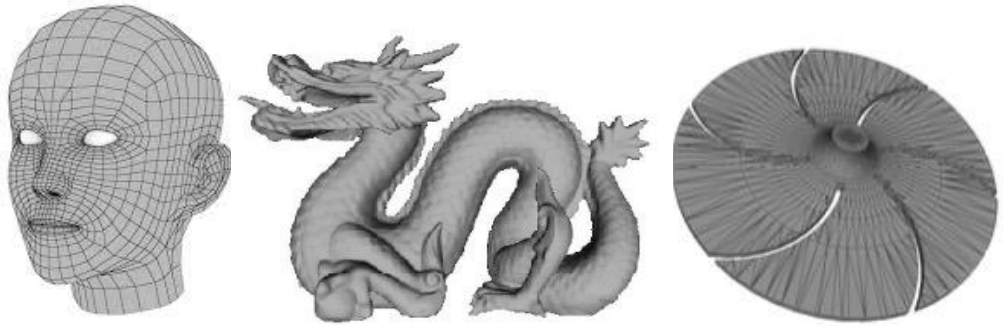
3D网格



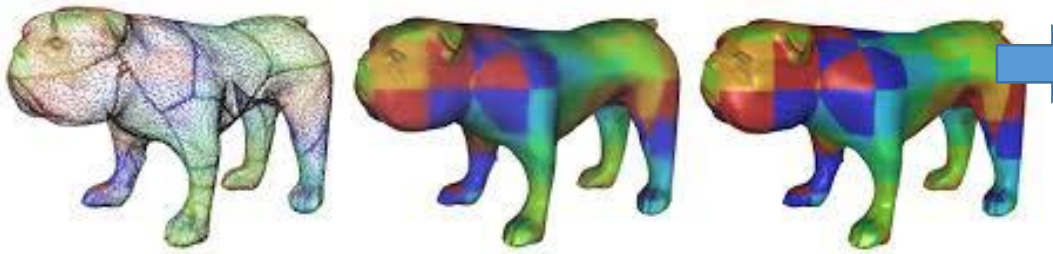
彩色纹理网格

深度图像

1 研究背景



3D网格

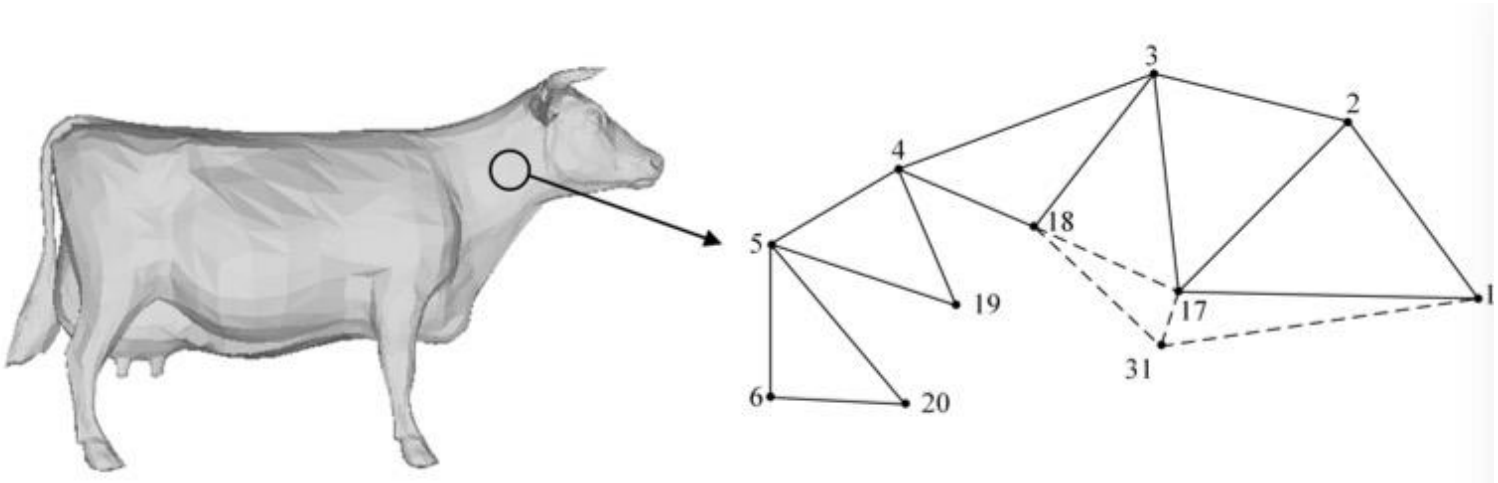


彩色纹理网格

深度图像

- 数字城市
- 3D渲染
- 文物保护
- 3D计算
- 逆向工程
- 虚拟现实(AR)
- 3D激光扫描
- 计算机辅助设计(CAD)
- ...

1 研究背景

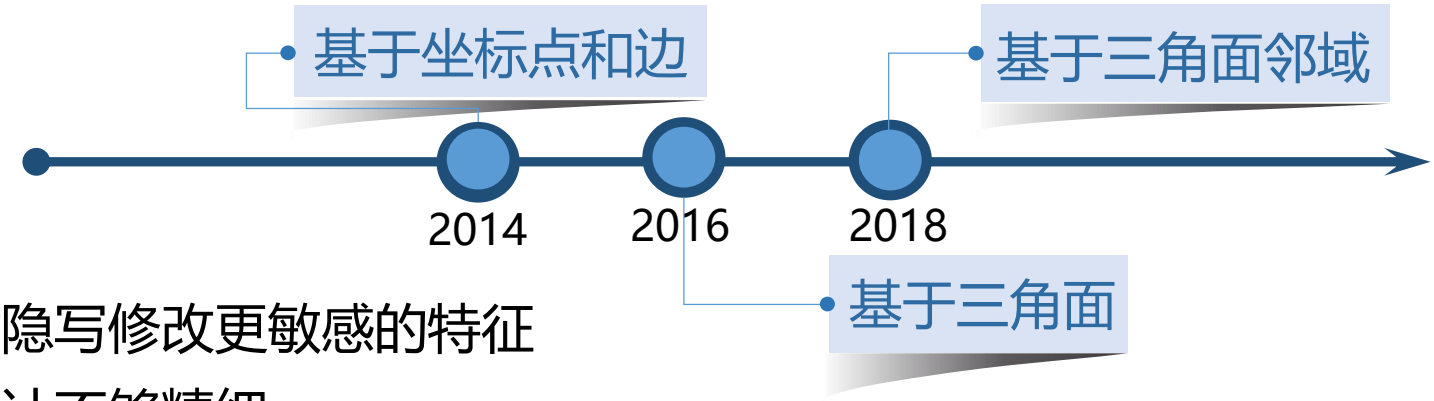


Vertex List				Face Information	
Index of vertex	x-axis coordinate	y-axis coordinate	z-axis coordinate	Index of face	elements in each face
1	$v_{1,x}$	$v_{1,y}$	$v_{1,z}$	1	(17,1,2)
2	$v_{2,x}$	$v_{2,y}$	$v_{2,z}$	2	(3,2,17)
3	$v_{3,x}$	$v_{3,y}$	$v_{3,z}$	3	(4,3,18)
4	$v_{4,x}$	$v_{4,y}$	$v_{4,z}$	4	(5,4,19)
5	$v_{5,x}$	$v_{5,y}$	$v_{5,z}$	5	(6,5,20)
6	$v_{6,x}$	$v_{6,y}$	$v_{6,z}$	...	...
...	...	...	...	16	(31,17,1)
17	$v_{17,x}$	$v_{17,y}$	$v_{17,z}$	17	(18,17,31)
18	$v_{18,x}$	$v_{18,y}$	$v_{18,z}$	...	...
19	$v_{19,x}$	$v_{19,y}$	$v_{19,z}$	241	(17,18,3)
20	$v_{20,x}$	$v_{20,y}$	$v_{20,z}$	...	...
...	...	...	...	...	...
31	$v_{31,x}$	$v_{31,y}$	$v_{31,z}$	...	...
...	...	...	...	...	...

1 研究背景

3D网格隐写分析与隐写的发展历程

3D网格隐写分析

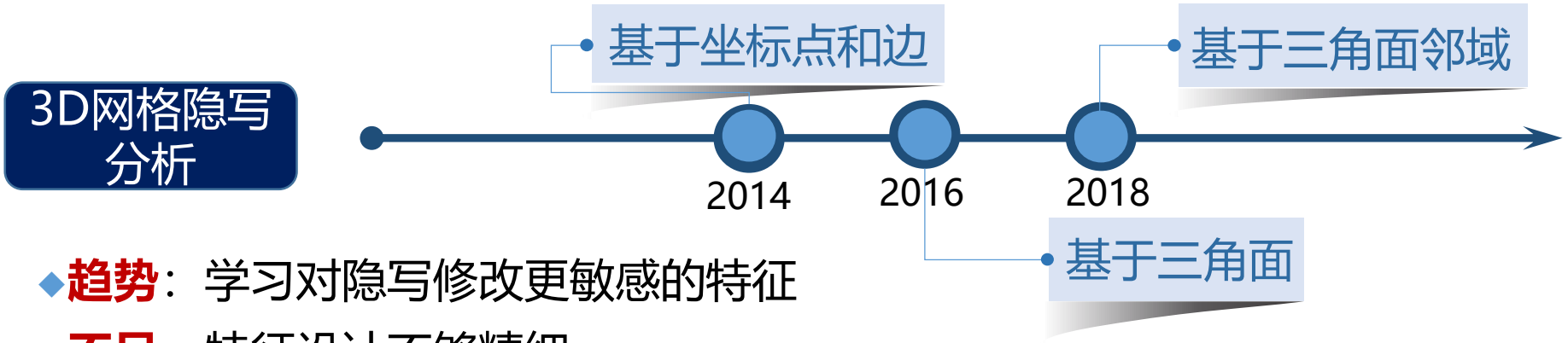


- **趋势**：学习对隐写修改更敏感的特征
- **不足**：特征设计不够精细

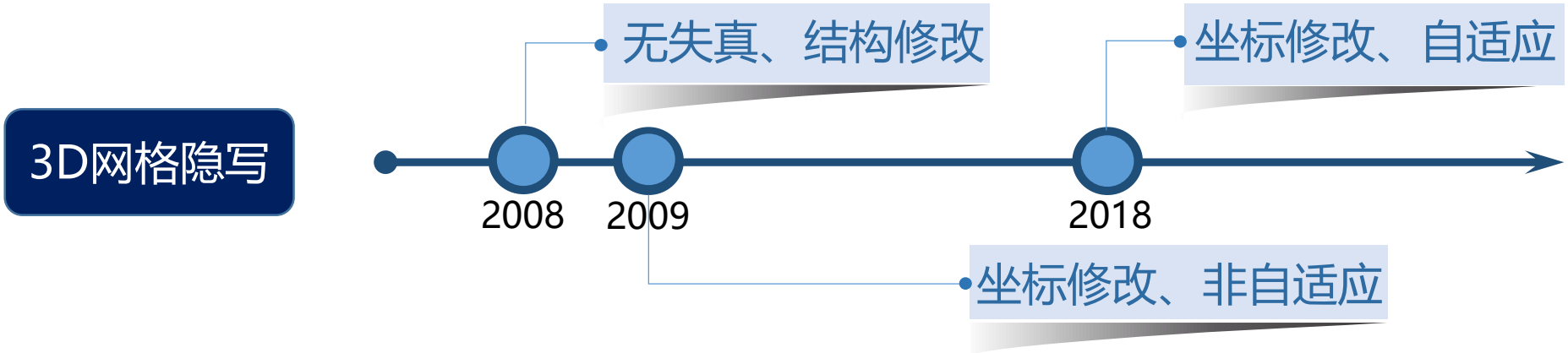
1 研究背景

中国科学技术大学 University of Science and Technology of China

3D网格隐写分析与隐写的发展历程



- **趋势**：学习对隐写修改更敏感的特征
- **不足**：特征设计不够精细

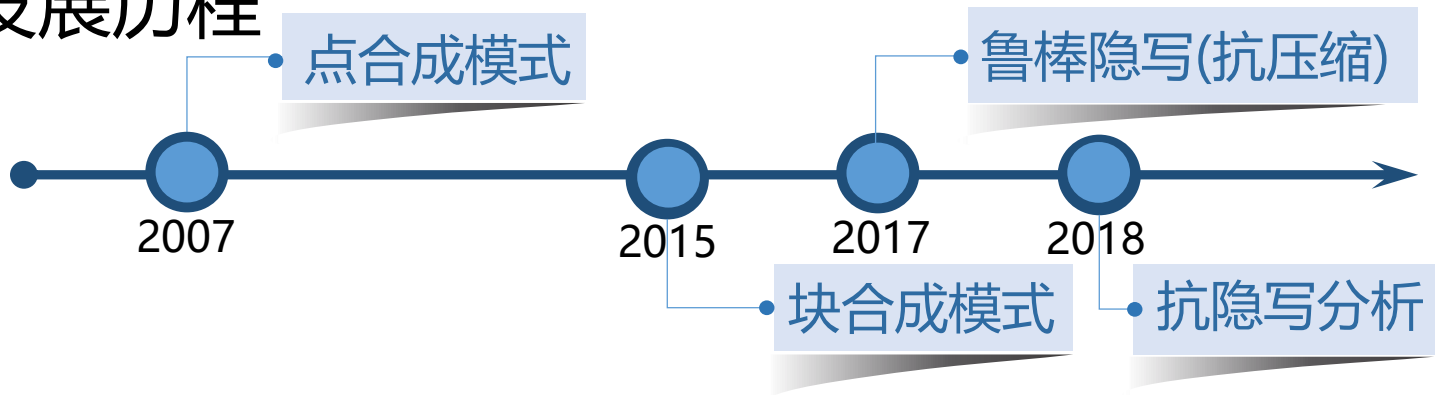


- **趋势**：从非自适应到自适应发展，寻找更适合的位置嵌入
- **不足**：缺乏自适应算法设计

1 研究背景

纹理隐写的发展历程

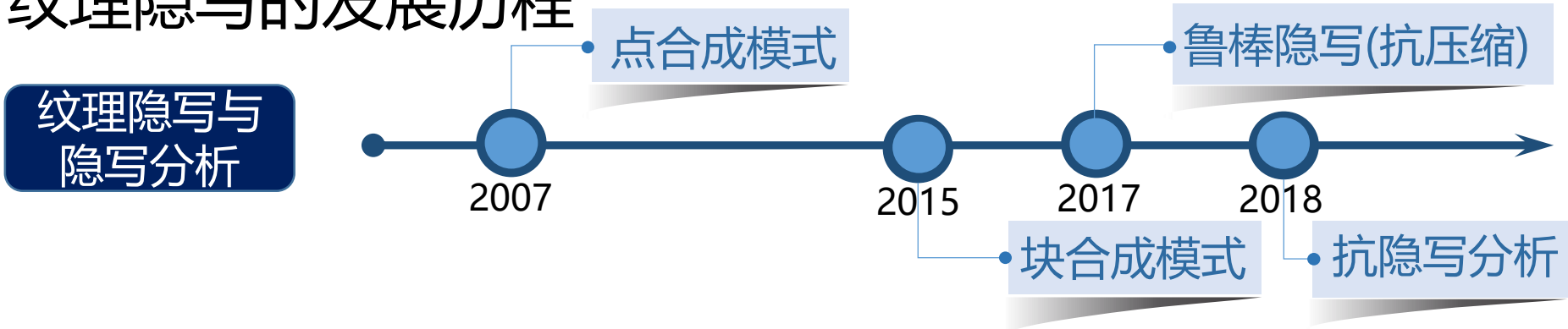
纹理隐写与隐写分析



- **趋势**：纹理合成向高视觉质量、大容量隐写方向发展
- **不足**：缺乏抗检测性

1 研究背景

纹理隐写的发展历程



- **趋势**：纹理合成向高视觉质量、大隐写容量方向发展
- **不足**：缺乏抗检测性

深度图像（RGBD）隐写的发展历程

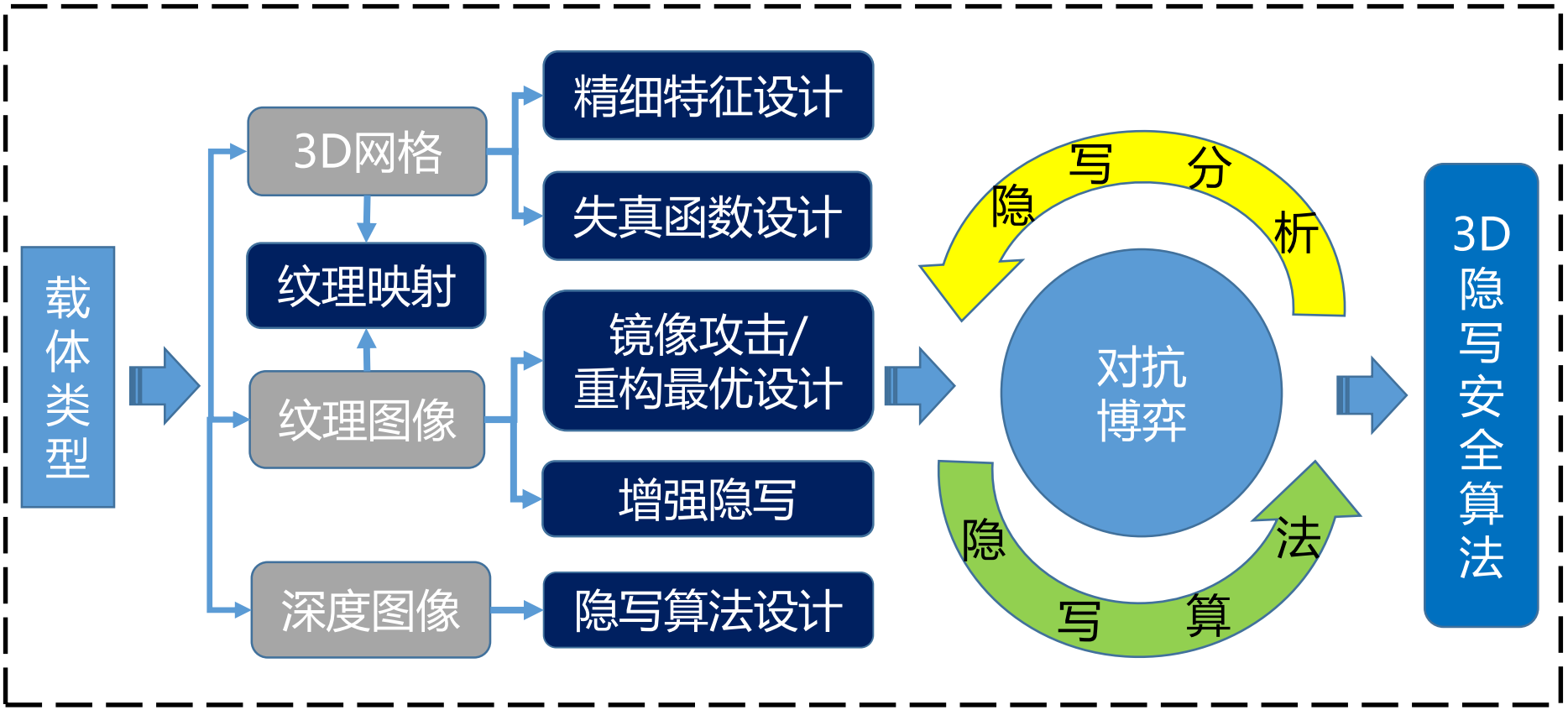
- **趋势**：尚未有RGBD图像隐写的研究

提纲

- 研究背景
- 研究内容
- 创新点
- 研究成果

2 研究内容

总体方案图


载体类型
3D网格
纹理映射
纹理图像
深度图像
精细特征设计
失真函数设计
镜像攻击/
重构最优设计
增强隐写
隐写算法设计
隐写分析
对抗博弈
隐写算法
3D隐写安全算法

研究内容

1. 3D网格模型隐写方法
   - 3D网格隐写安全性分析
   - 基于最小化失真框架的3D网格模型隐写
2. 3D纹理合成隐写方法
   - 纹理图像合成隐写安全性分析
   - 基于3D纹理贴图的隐写方法
3. 3D深度图像隐写方法
   - 深度图像隐写算法

2 研究内容

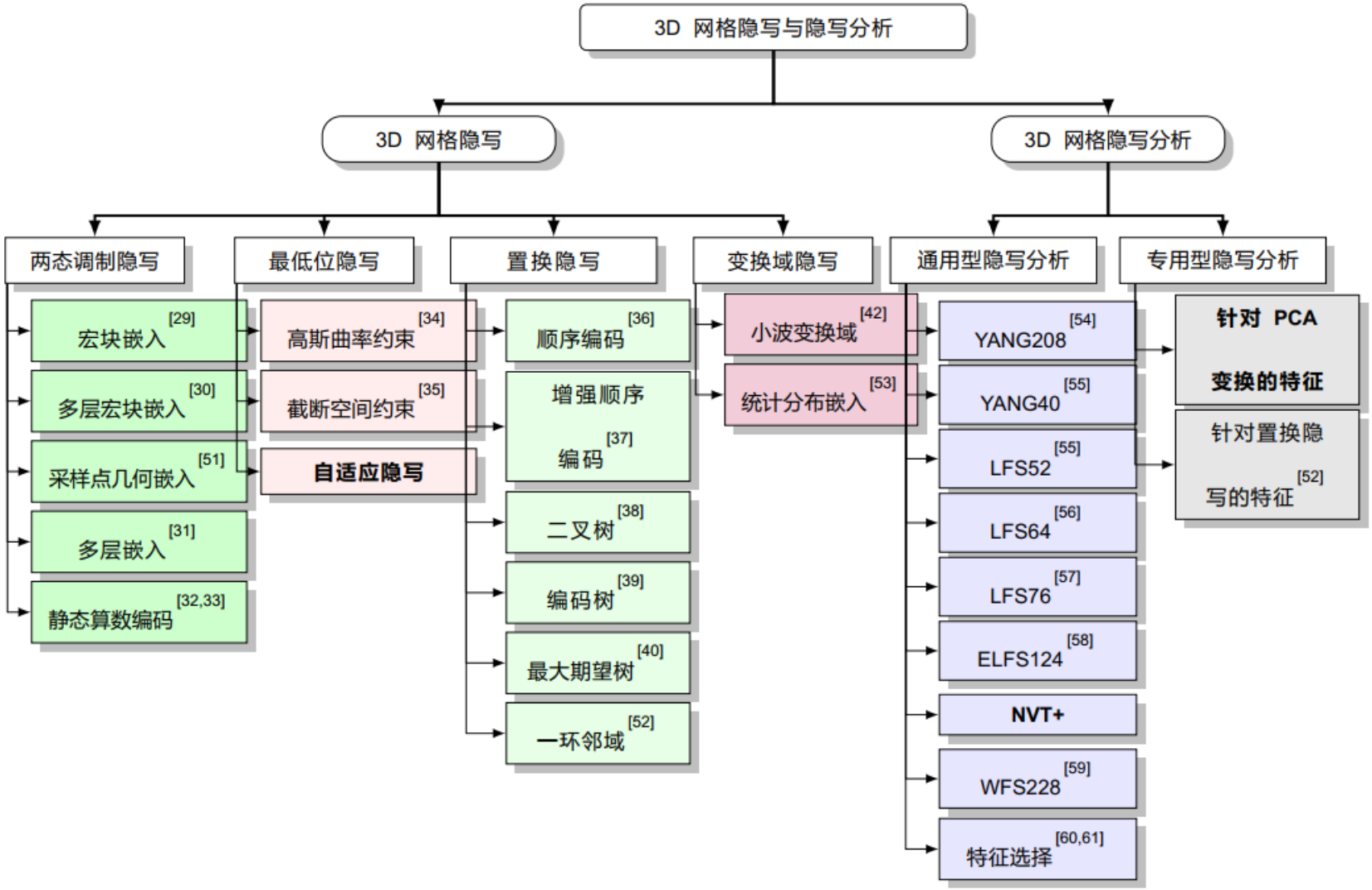
研究点1：

3D网格模型隐写方法

1. 3D网格隐写安全性分析
2. 基于最小化失真框架的3D网格模型隐写

2 研究内容

- 3D网格隐写与隐写分析分类系统



2 研究内容

○ 3D网格隐写安全性分析

- 研究目的

  现有3D隐写分析：基于**坐标点**信息或**边**信息的相关性强弱设计特征，隐写分析效果较差。

- 研究动机

  特点：3D网格模型隐写会破坏坐标点**邻域相关性**。猜想，是否存在一种代表邻域的元素，能够更敏感地察觉3D坐标点隐写扰动造成的影响？

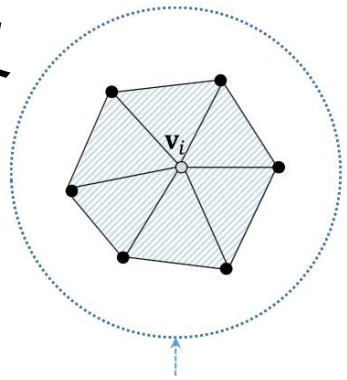
  提出的算法：由于坐标点或三角面涵盖区域较小，考虑**三角面一环邻域区域**作为单位元素，分析邻域相关性。

2 研究内容

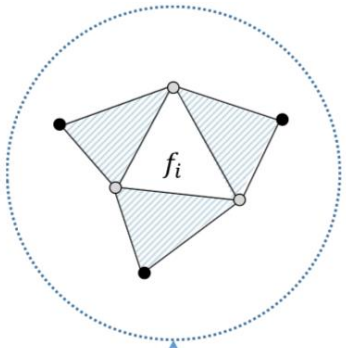
○ 3D网格隐写安全性分析

- 隐写分析框架

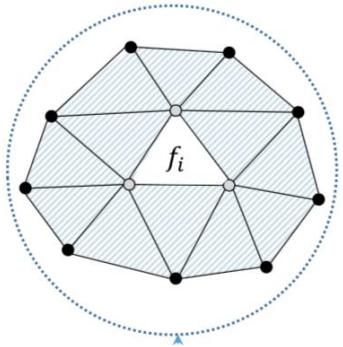
- 邻域定义



顶点的面邻域



三角面边相
邻的面邻域



三角面点相
邻的面邻域

2 研究内容

○ 3D网格隐写安全性分析

▪ 特征设计

1. 三角面法向量作为单元，抽取邻域特征（法向量投票张量）：

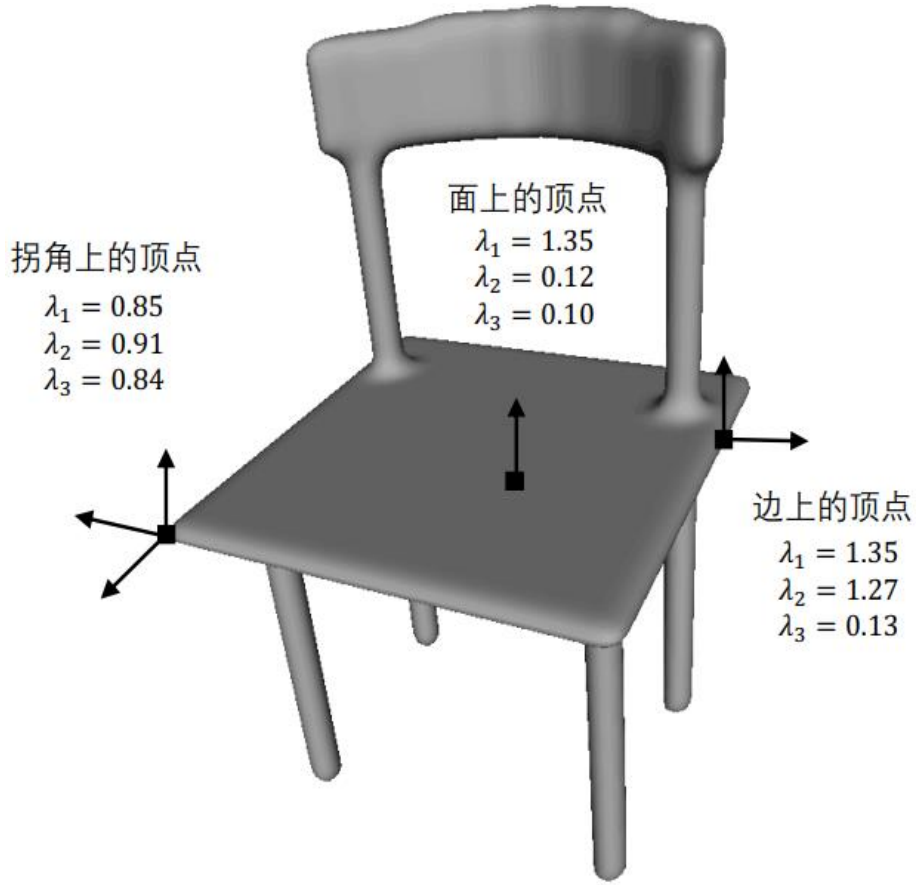
$$T_i = \sum_{j \in \Omega(i)} \mu_{ij} \mathbf{n}_j \cdot \mathbf{n}_j^T$$

2. 对张量进行特征值分解：

$$T_i = \sum_{k=1}^{3} \lambda_k e_k e_k^T$$

3. 构建残差特征：

$$\phi_k(i) = |\lambda_k(i) - \lambda_k'(i)|$$



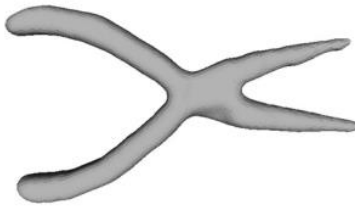
2 研究内容

○ 3D网格隐写安全性分析

▪ 实验评价方法

- 数据集：普林斯顿分割数据集 (PSB) [260/94]
  普林斯顿网格 (PMN) [6155/6155]
- 分类器：FLD集成分类器
  监督训练隐写分析器
- 检测错误率：

$$P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$$

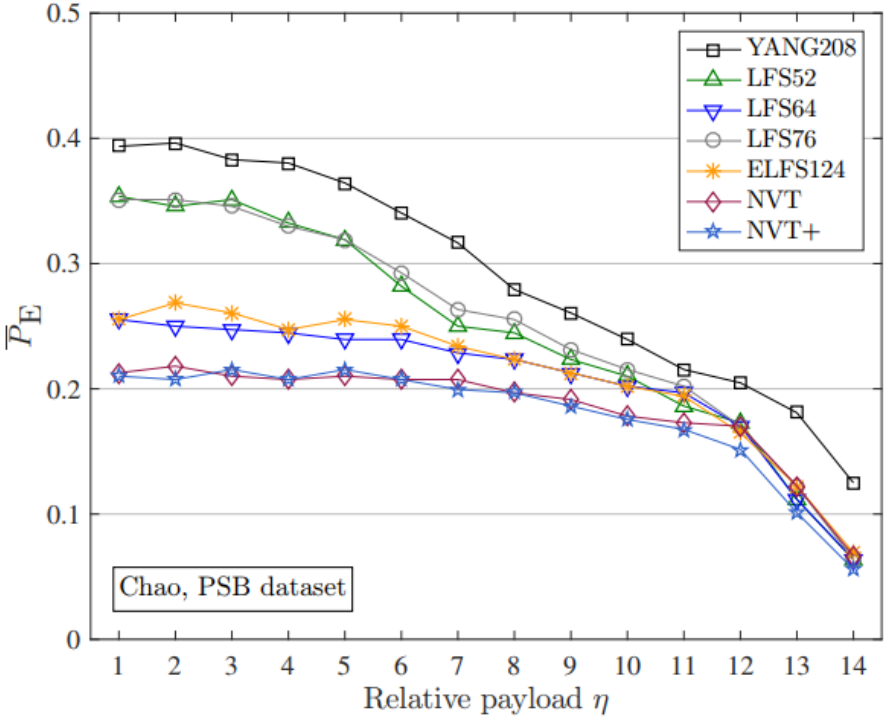


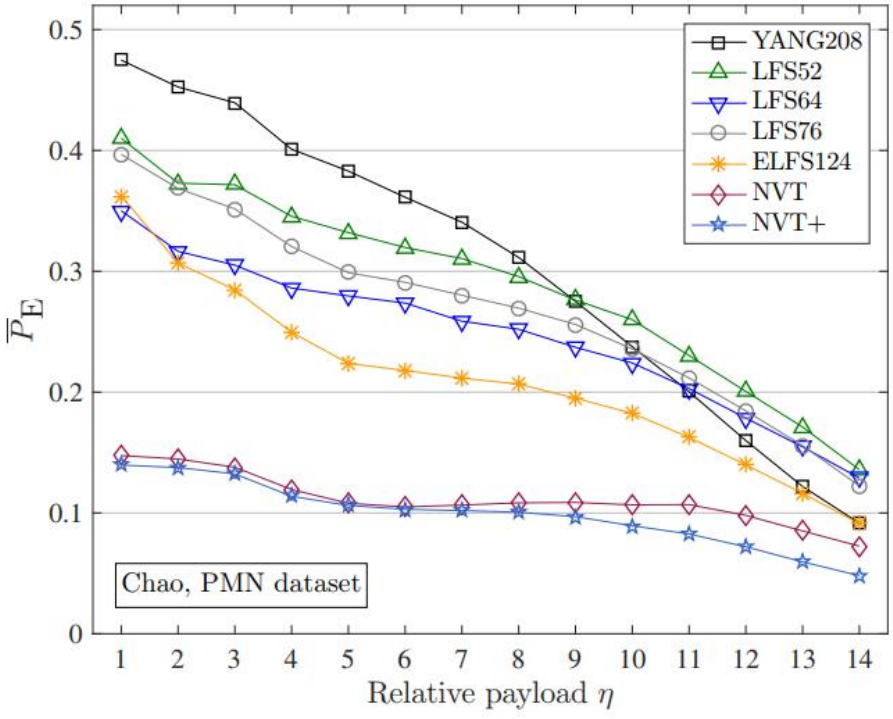
2 研究内容

○ 3D网格隐写安全性分析

▪ 实验结果

- 检测Chao隐写算法





NVT+特征的隐写分析检测率明显更低，表现更佳。

2 研究内容

○ 3D网格隐写安全性分析

▪ 结论

提出了基于**三角面邻域法向量张量投票模型**的特征来提升3D隐写分析性能，**形成了新的3D网格隐写安全评测方法**。

该研究成果，已发表在3D计算机图形可视化领域国际期刊TVCG上。

2 研究内容

○ 基于最小化失真框架的3D网格模型隐写

- 研究目的

  现有3D隐写：基本上通过**坐标点空域调制嵌入秘密消息**，未考虑**拓扑关系**的约束。

- 研究动机

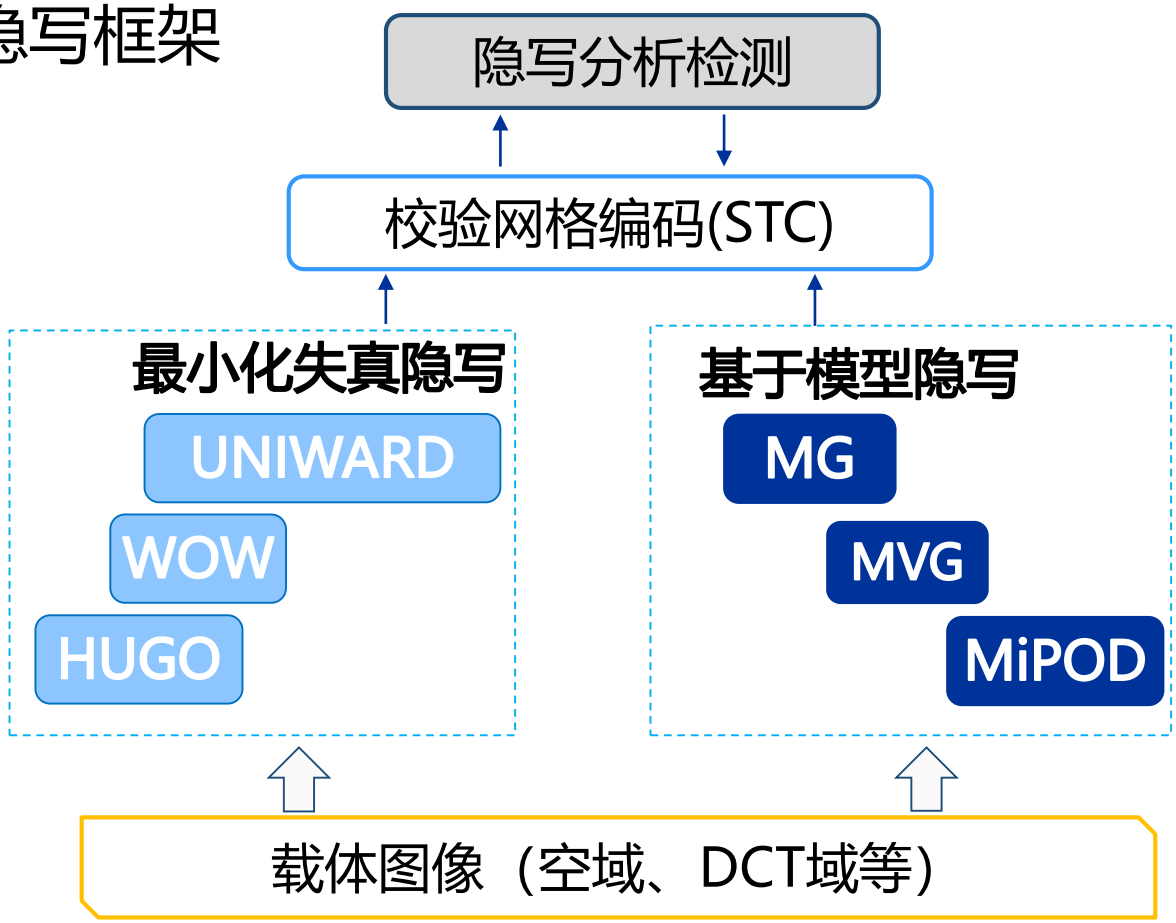
  特点：图像上，由于修改不同的像素值对隐写分析的检测的影响程度是不同的，因此需要采用**自适应隐写**算法进行消息嵌入。

  提出的算法：类比至3D网格隐写，不同坐标点应赋予不同的修改权重。

2 研究内容

○ 基于最小化失真框架的3D网格模型隐写

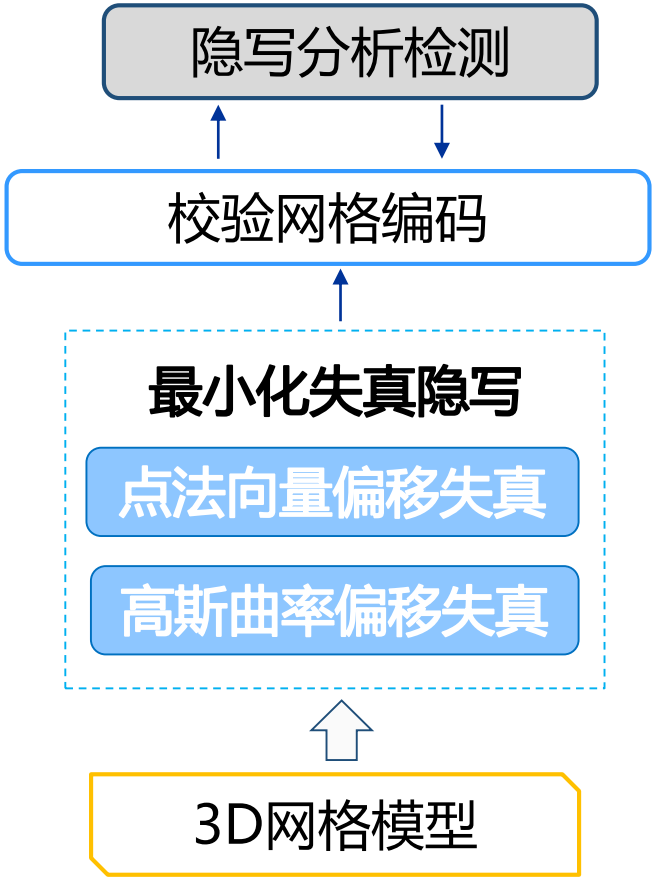
- 图像自适应隐写框架



2 研究内容

○ 基于最小化失真框架的3D网格模型隐写

- 3D网格自适应隐写框架

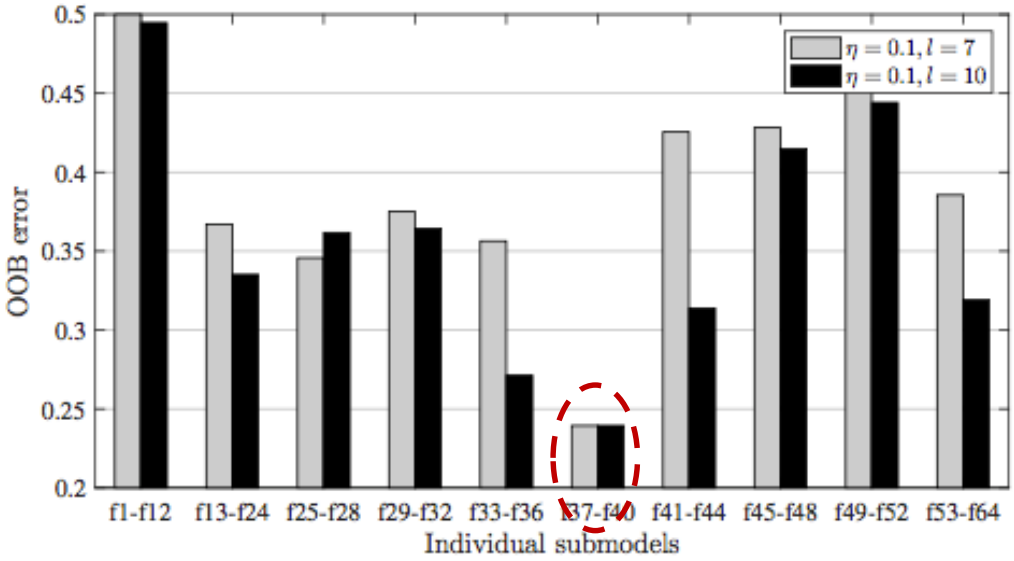


2 研究内容

○ 基于最小化失真框架的3D网格模型隐写

- 失真函数设计（方案一）

实验发现，基于**顶点法向量**的隐写分析特征(f37-f40)对载体和载密的分类性能最好，因此通过**约束法向量敏感的坐标点**，能最大限度提高隐写算法抗检测性能。



不同隐写分析子特征的分类错误率

2 研究内容

○ 基于最小化失真框架的3D网格模型隐写

▪ 失真函数设计（方案一）

实验发现，基于**顶点法向量**的隐写分析特征(f37-f40)对载体和载密的分类性能最好，因此通过**约束法向量敏感的坐标点**，能最大限度提高隐写算法抗检测性能。

失真函数：顶点法向量偏移量的倒数

$$\rho_i = \frac{1}{\ln(||\overrightarrow{N}_{\mathbf{v}_i} - \overrightarrow{N}_{\mathbf{v}_i'}||_2 + 1) + \sigma}. \qquad i = 1, 2, \ldots, N$$

顶点法向量经过平滑前后的偏移量越小，失真越大，越不修改它。

2 研究内容

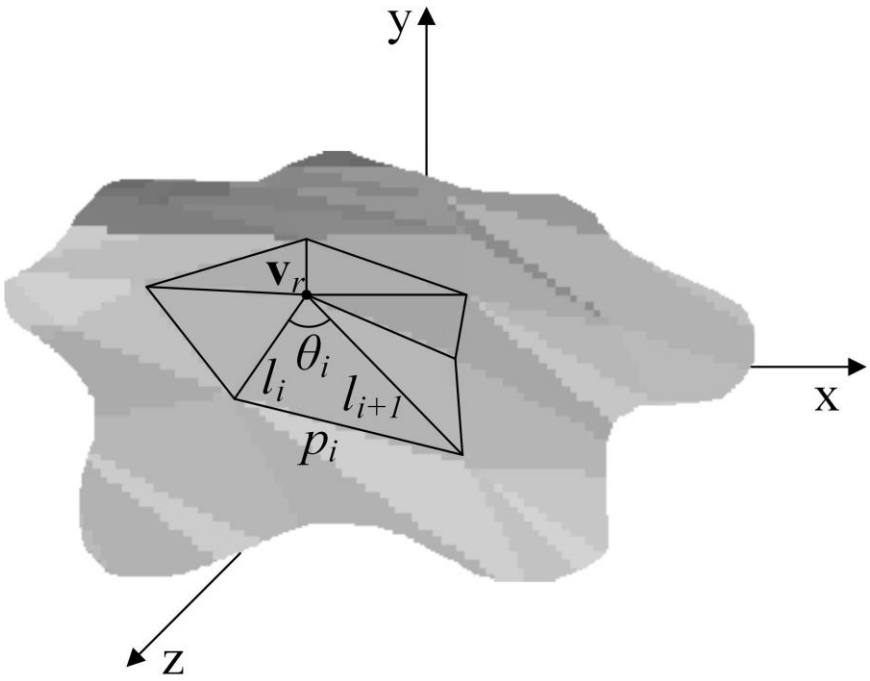
○ 基于最小化失真框架的3D网格模型隐写

▪ 失真函数设计（方案二）

规则：复杂区域定义较小的失真。
顶点越尖锐，复杂度越高，可通过高斯曲率进行评价。

失真函数：高斯曲率的倒数

$$\rho_i' = \frac{1}{|K(\mathbf{v}_i)|^{\alpha} + \sigma}, \qquad 1 \le i \le N$$



某一顶点的局部区域

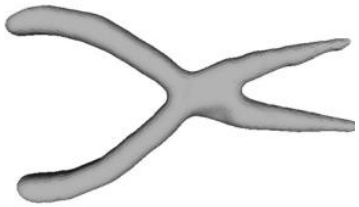
2 研究内容

○ 基于最小化失真框架的3D网格模型隐写

▪ 实验评价方法

- 数据集：普林斯顿分割数据集 (PSB) [260/94]
  普林斯顿网格 (PMN) [6155/6155]
- 分类器：FLD集成分类器
  监督训练隐写分析器
- 检测错误率：

$$P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$$

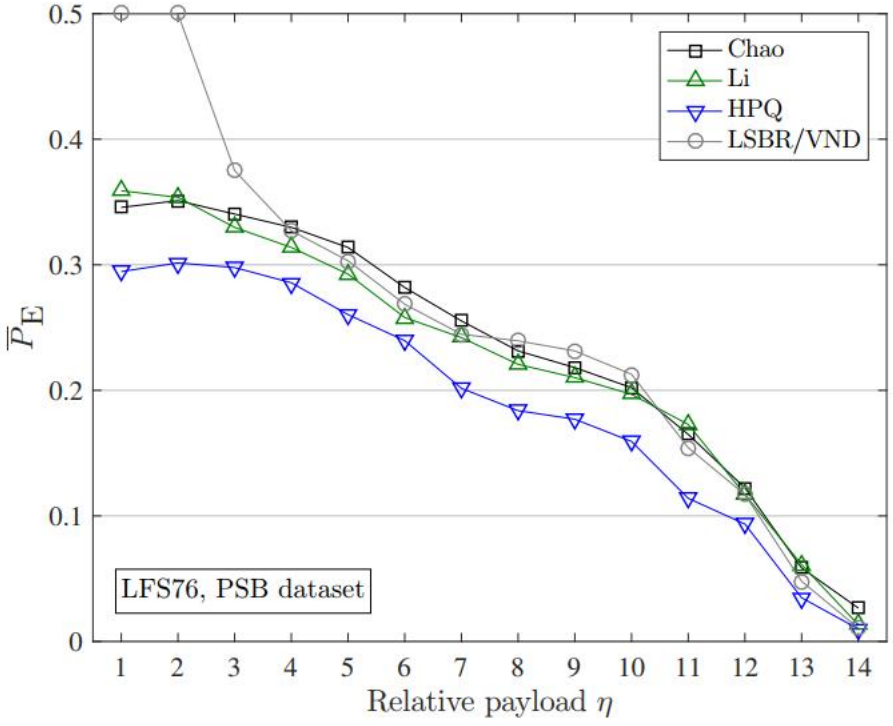


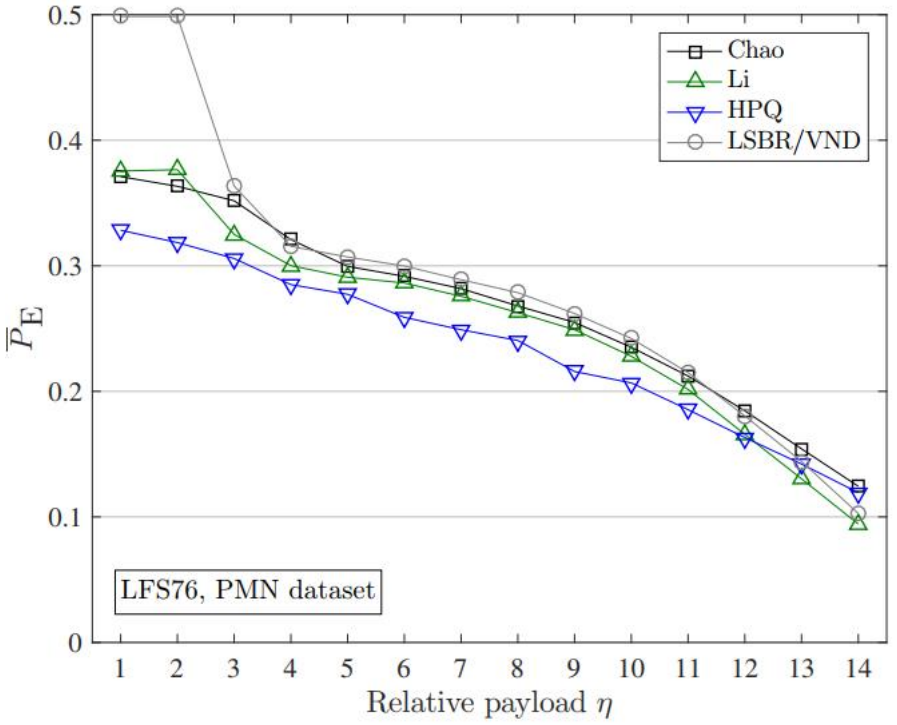
2 研究内容

○ 基于最小化失真框架的3D网格模型隐写

▪ 实验结果

- 采用LFS76检测隐写算法的安全性





VND隐写算法的安全性在低嵌入率下的检测错误率较高，表现更佳。

2 研究内容

○ 基于最小化失真框架的3D网格模型隐写

▪ 结论

分析了不同子隐写分析特征对隐写扰动的检测能力，提出了基于**坐标点法向量**的失真函数，有效提升了3D网格载密模型的安全性。

该研究成果，已发表在多媒体国际期刊TMM上。

2 研究内容

研究点2：

3D纹理合成隐写方法

1. 纹理图像合成隐写安全性分析
2. 基于3D纹理贴图的隐写方法

2 研究内容

○ 纹理图像合成隐写安全性分析

- 研究目的

  现有算法：由于纹理图像的**全局复杂性**和**近似周期性，**传统的隐写分析方法难以对现有纹理隐写算法进行检测，这对评估隐写的安全性带来了极大的挑战。

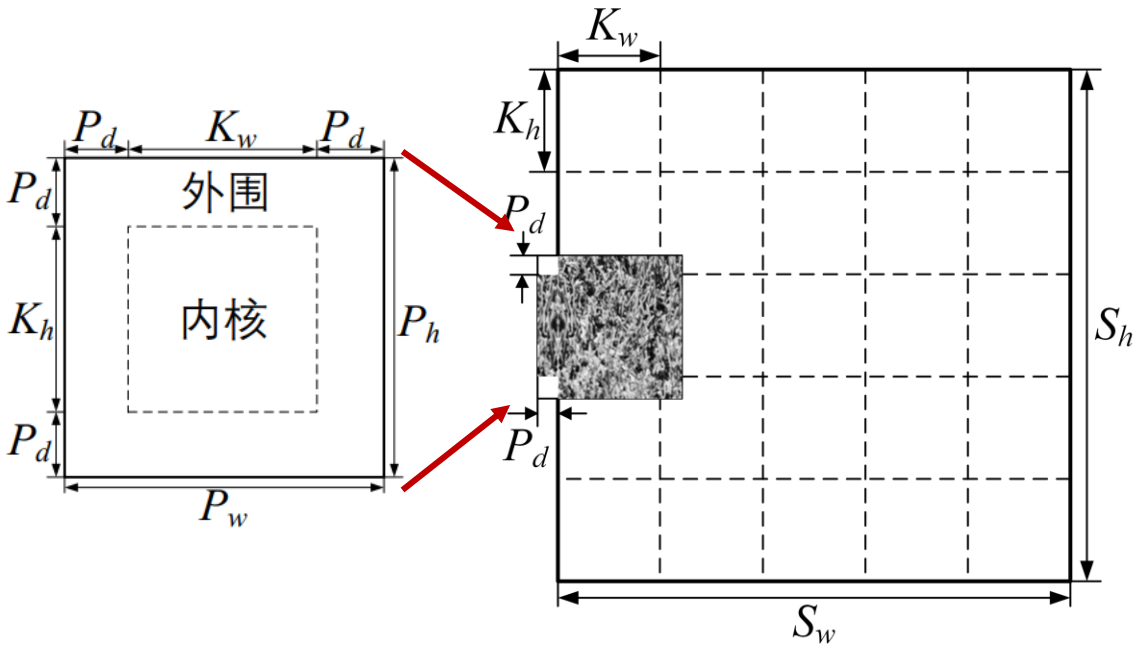
- 研究动机

  提出的算法：已有的纹理合成隐写方法有**漏洞**。攻击者通过分析合成图像中的每一个图像单元来判断该图像块是否来自于原始**图像边缘重构**图像边界，迭代多次能够重构出原图。

2 研究内容

○ 纹理图像合成隐写安全性分析

▪ 纹理图像隐写算法

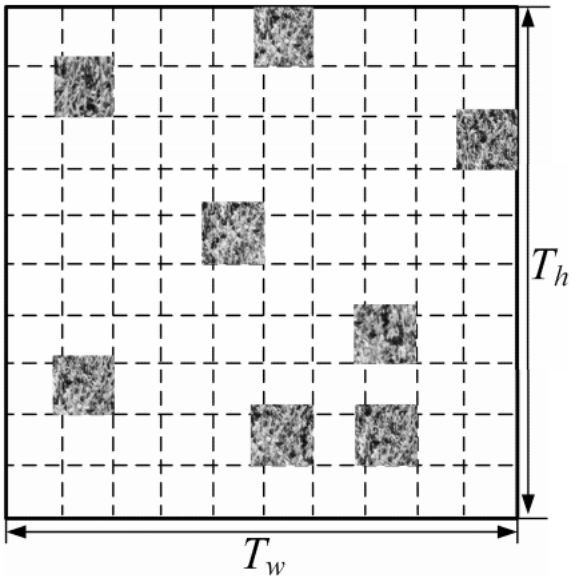


纹理单元　　生成候选图像块的过程

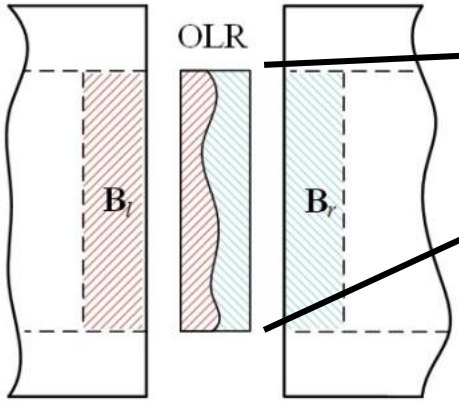
2 研究内容

○ 纹理图像合成隐写安全性分析

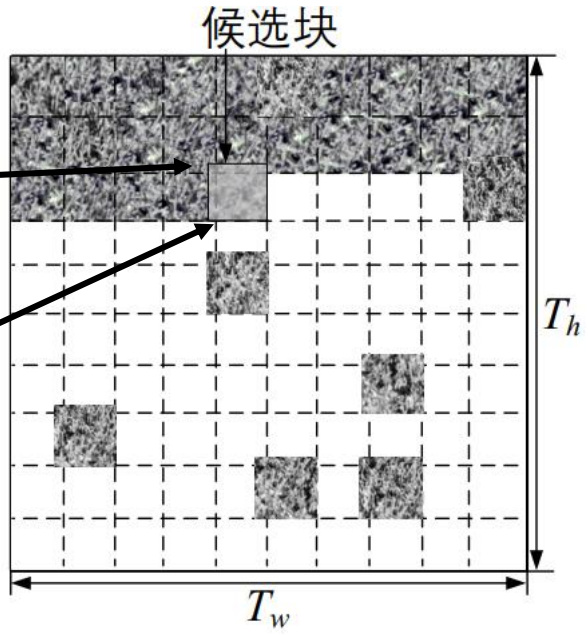
- 纹理图像隐写算法



空白的合成图像



纹理合成示意图



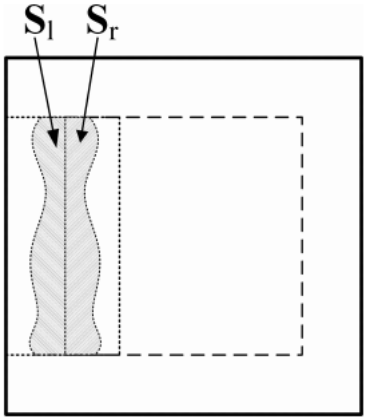
生成合成图像的过程

候选图像单元与当前待合成图像单元计算MSE，并排序。当前消息比特编码值对应的图像单元作为待合成图像单元。

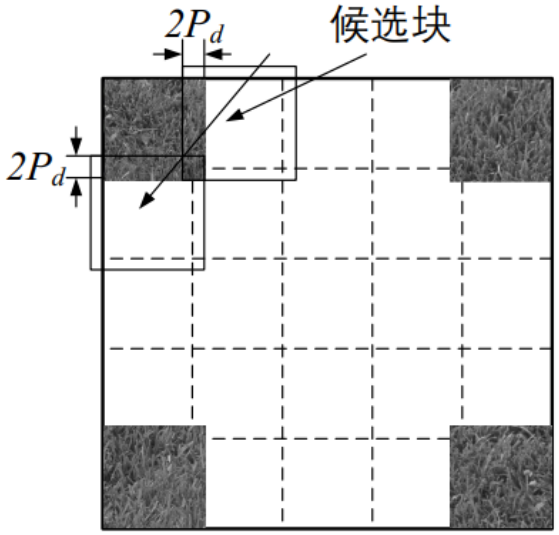
2 研究内容

○ 纹理图像合成隐写安全性分析

- 基于图像边界镜像延拓检测（方案一）



图像边界呈镜像对称



图像重构顺序（先四个角）

通过比较核区域与边界区域镜像对称的程度重构原图，逐步重构外围，直至全图重构完毕。接着，在隐写过程中与合成图进行匹配，进行消息提取。

2 研究内容

○ 纹理图像合成隐写安全性分析

▪ 实验评价方法

- 数据集：Brodatz 纹理图像库 [112]
- 消息正确提取率

▪ 实验结果

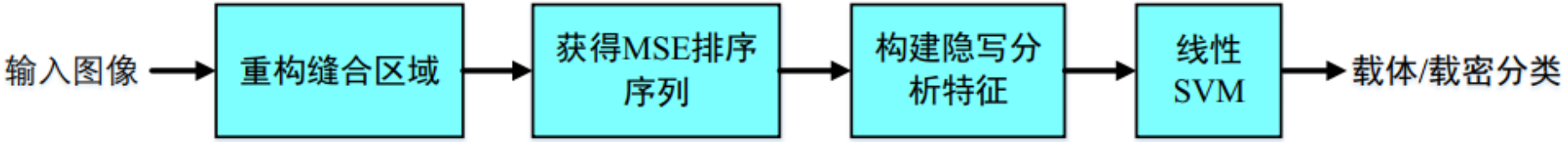
图像正确重构率	漏警率 P_{MA}	虚警率 P_{FA}	消息正确提取率
96.77%	0	5.71%	94.66%

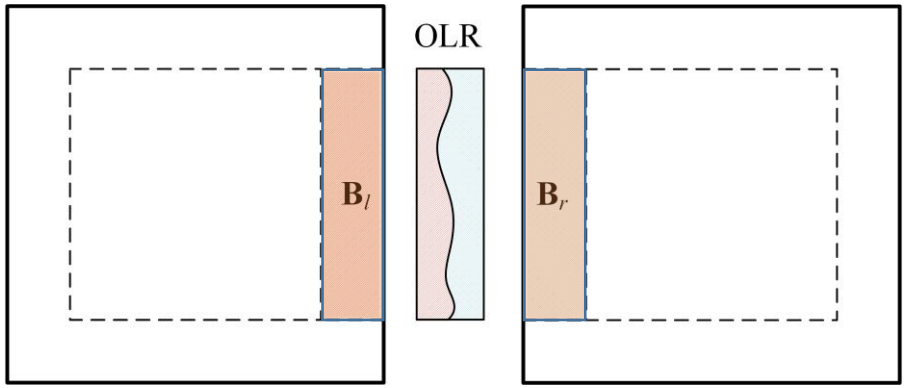
少数情况下无法正确提取消息，有两种可能：其一是重构过程出错；其二是即使原图重构正确，但是重构的某些图像单元来源于候选图像单元。

2 研究内容

○ 纹理图像合成隐写安全性分析

- 统计最优性特征构造（方案二）





待合成的两个纹理单元

对于任一重叠区域，重构B_l和B_r区域，重新合成，检测匹配的最优性程度。

2 研究内容

○ 纹理图像合成隐写安全性分析

▪ 隐写分析特征设计

记$R = \{r_i | r = 1, 2, ..., N\}$为载体或载密图像的最优性排序值集合。均值、中值、方差和峰度作为隐写分析特征。

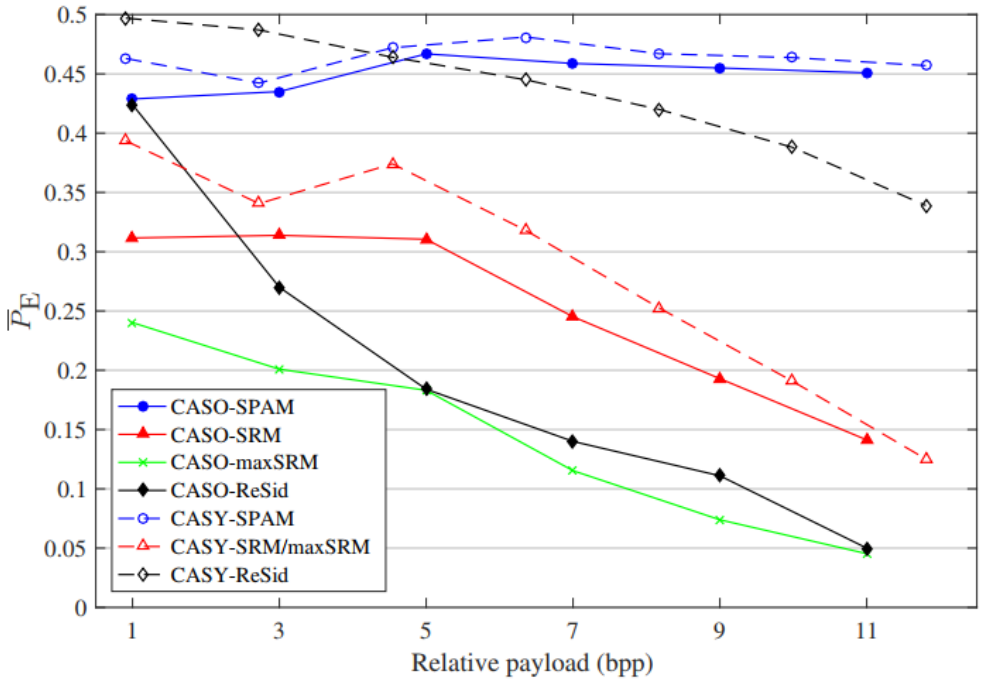
▪ 实验评价方法

- 数据集：Brodatz 纹理图像库（经过数据增强）
- 分类器：SVM分类器

  监督训练隐写分析器
- 检测错误率

2 研究内容

- **纹理图像合成隐写安全性分析**
  - 实验结果
    - CASO为纹理隐写算法。针对CASO的隐写分析性能比较。



相比于其他隐写分析算法（SPAM/SRM），ReSid隐写分析算法具备更低的检测错误率，表现更佳。

2 研究内容

○ 纹理图像合成隐写安全性分析

▪ 结论

分析了纹理合成隐写算法的漏洞，提出了一种基于**镜像重构**的**攻击方法**，能够重构出原始纹理图像，并且能够高概率提取嵌入的信息。

进一步，提出了另一种基于**重构区域最优性**的隐写分析方法，形成了新的纹理图像隐写安全评测方法。

该研究成果，已发表在图像处理国际期刊TIP和JVCIR上。

2 研究内容

○ 基于3D纹理贴图的隐写方法

- 研究目的

  现有算法：纹理合成隐写算法安全性较低。

- 研究动机

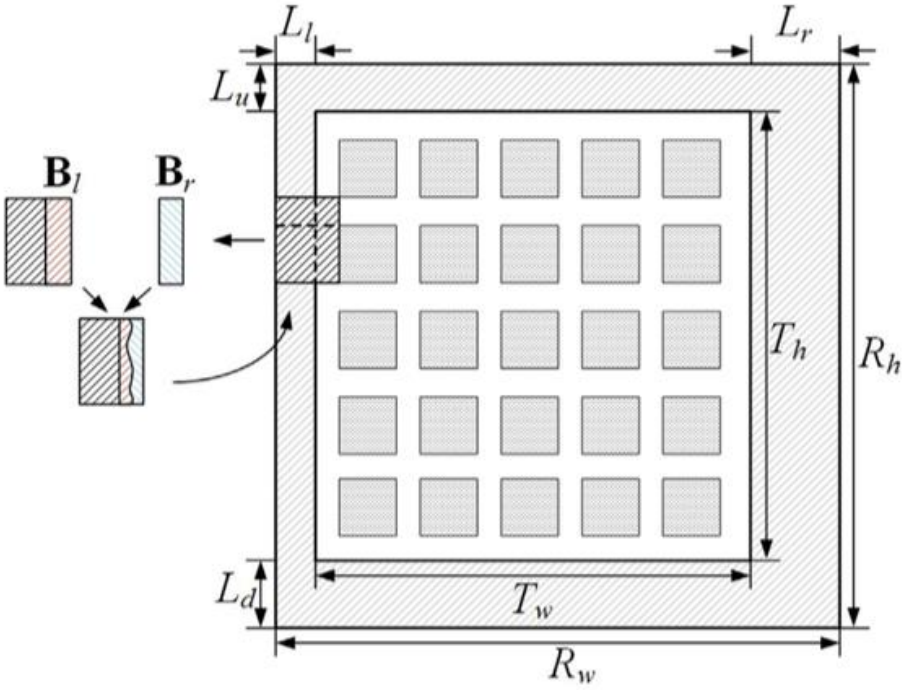
  提出的算法：通过对边界区域的填充，使得攻击者难以估计合成块大小。结合纹理隐写和3D网格隐写，基于MeshLab平台进行纹理贴图，实现多域联合隐写。

2 研究内容

○ 基于3D纹理贴图的隐写方法

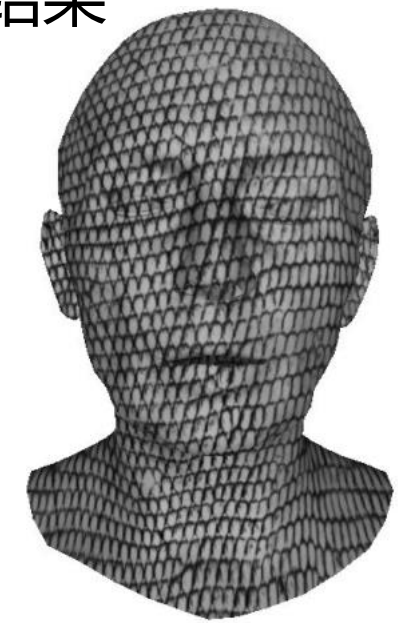
▪ 方案设计

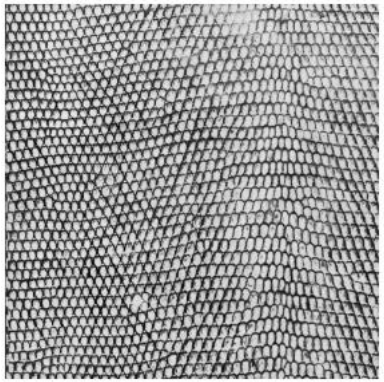
通过边界填充，使攻击者难以估计合成纹理单元的大小，进而难以进行隐写分析。



2 研究内容

- **基于3D纹理贴图的隐写方法**
  - 实验结果



2 研究内容

○ 基于3D纹理贴图的隐写方法

▪ 结论

设计了基于**图像边界纹理块扰动**的**增强纹理图像隐写**算法，提高了纹理隐写算法的抗检测性。同时，实现了实际有效的**载密图像3D纹理贴图**，达到了多域隐写的目的。

该研究成果，已发表在图像处理国际期刊JVCIR上。

2 研究内容

研究点3：

3D深度图像隐写方法

1. 深度图像隐写算法

2 研究内容

○ 深度图像隐写算法

- 研究目的

  多模态集成通信是隐写术的一种特殊应用，例如彩色图像灰度化和深度图像彩色化等高维数据低维化过程，并且需要保证载体的可自重构性。

  深度图像通常存储为同一路径下的两个单独文件，在使用中必须同时传输或加载，而且存在深度文件容易丢失的问题。

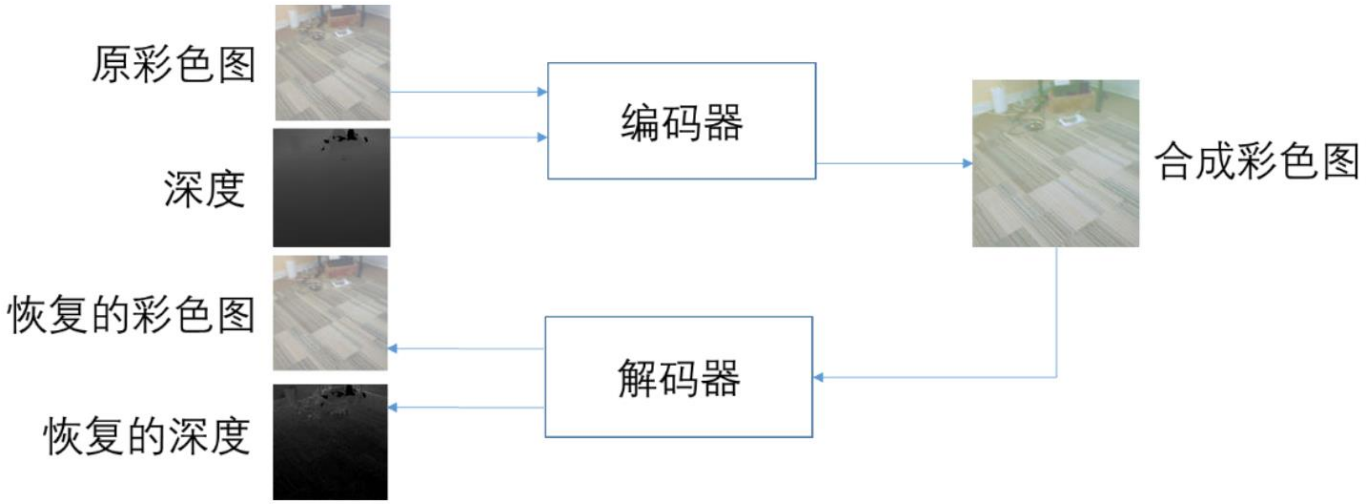
- 研究动机

  通过**图像合成和图像解码**来实现隐写和消息提取。

2 研究内容

○ 深度图像隐写算法

▪ 算法框架



隐写过程：

$$I_e = \mathbb{E}([\mathcal{E}_c(I_o); \mathbb{E}_d(D_o)])$$

图像提取过程：

$$[I_d; D_d] = \mathbb{D}(I_e).$$

2 研究内容

○ 深度图像隐写算法

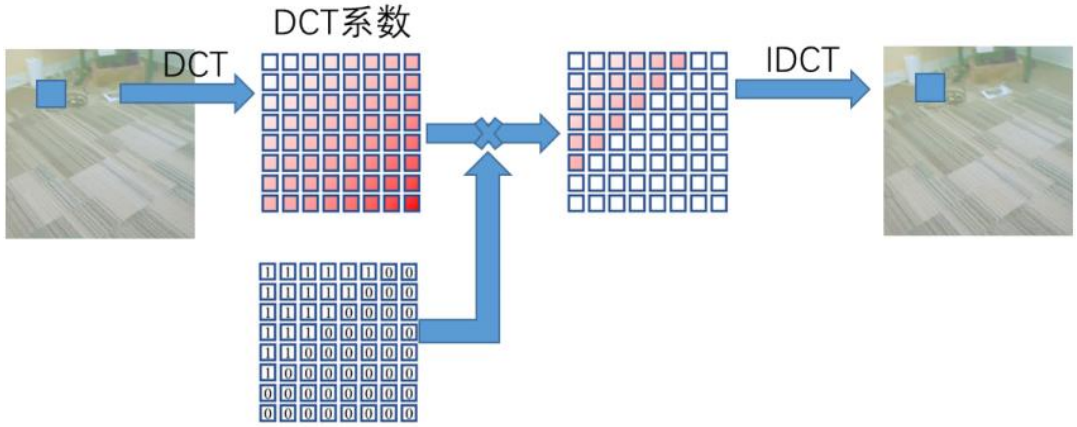
▪ 网络结构

图像隐藏网络：特征编码器：1X卷积层+2X残差模块+2X下采样模块
特征解码器：2X上采样模块+2X残差模块+1X卷积层

图像提取网络：1X卷积层+8X残差模块+2X卷积层

对抗生成网络：采用PatchGAN网络结构

JPEG压缩模拟网络：步径为8的8X8卷积核，采用门函数矩阵激活。



JPEG压缩模拟网络

2 研究内容

○ 深度图像隐写算法

▪ 目标损失函数

1. 生成网络的损失函数
   - 均方误差损失：$\mathcal{L}_I(I_o, I_e) = \|I_o - I_e\|^2$

     $\mathcal{L}_I(I_o, I_d, D_o, D_d) = \|I_o - I_d\|^2 + \|D_o - D_d\|^2$
   - 一致损失：$\mathcal{L}_{conf}(I_o, I_e) = \|\max(|I_o - I_e| - \tau, 0)\|_1$
   - 对比度损失：$\mathcal{L}_{cont}(I_o, I_e) = \|VGG_l(I_o) - VGG_l(I_e)\|^2$
   - 局部结构损失：$\mathcal{L}_{str}(I_o, I_e) = \|Var(I_o) - Var(I_e)\|_1$
   - 对抗损失：$\mathcal{L}_{gen} = \mathbb{E}[\log_{x \in I_o}(\mathcal{R}(x))]$
   - **总损失**：$\mathcal{L}_G = \mathcal{L}_I(I_o, I_e) + \mathcal{L}_I(I_o, I_e, D_o, D_d) + \alpha\mathcal{L}_{conf} + \beta\mathcal{L}_{cont} + \gamma\mathcal{L}_{str} + \eta\mathcal{L}_{gen}$
2. 判决网络的损失函数
   - 对抗损失：$\mathcal{L}_{dis} = \mathbb{E}[\log_{x \in I_o}(\mathcal{R}(x))] - \mathbb{E}[\log_{y \in I_d}(1 - \mathcal{R}(y))]$
   - **总损失**：$\mathcal{L}_D = \mathcal{L}_{dis}$

2 研究内容

○ 深度图像隐写算法

▪ 实验评价方法

- 数据集：NYU 深度数据集 V2

- 评价指标：峰值信噪比（PSNR）

2 研究内容

○ 深度图像隐写算法

▪ 无损隐写质量评估

1. 定量评估

表 5.1 不抗 JPEG 压缩的隐写前后图像 PSNR 值

隐写图像 I_o	提取的 RGB 图像 I_d	提取的深度 D_d
42.45	41.56	47.16

2. 定性评估

原RGB图

原深度图

重构RGB图

重构深度图

2 研究内容

○ 深度图像隐写算法

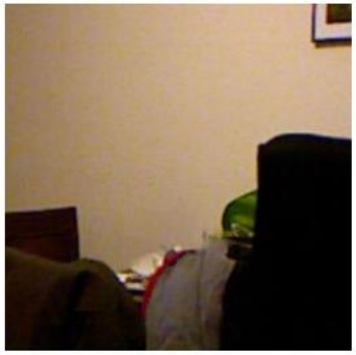
▪ 鲁棒隐写质量评估

1. 定量评估

表 5.2 抗 JPEG 压缩的隐写前后图像 PSNR 值

载密 I_o	提取的深度 D_d				提取的 RGB 图像 I_d		
	模拟压缩	QF=75	QF=95	QF=100	QF=75	QF=95	QF=100
18.52	38.02	31.64	35.02	35.02	27.12	28.77	29.01

2. 定性评估（经QF=95压缩）


原RGB图


原深度图


重构RGB图


重构深度图

2 研究内容

○ 深度图像隐写算法

▪ 结论

提出了基于**深度图像合成**的隐写算法，该方法作为多模态集成通信的一种特殊应用，能够实现高维信息的自重构，达到隐蔽通信的用途。

在网络设计上，本文采用深度卷积网络进行隐写和信息提取，并且采用对抗生成网络保证载体和载密的不可区分性。

该研究成果结合3D渲染实现高质量立体照片，即将成文投稿。

提纲

- 研究背景
- 研究内容
- 创新点
- 研究成果

3 创新点

总结

1. 提出了基于三角面邻域法向量张量投票模型的特征来提升3D隐写分析性能，形成了新的3D网格隐写安全评测方法。
2. 提出了基于坐标点的失真函数，并采用校验网格编码实施自适应3D网格隐写，有效提升了3D网格载密模型的安全性。
3. 提出了一种基于镜像重构的纹理图像隐写攻击方法，以及一种基于图像单元边界缝合最优性的统计特征以进行纹理图像隐写分析，形成了新的纹理图像隐写安全评测方法。
4. 提出了一种增强的纹理图像隐写算法，并将纹理图像隐写与3D网格模型结合，设计了3D纹理贴图隐写方法，实现了多域联合隐写。
5. 提出了基于图像合成的深度图像隐写算法。

提纲

- 研究背景
- 研究内容
- 创新点
- 研究成果

4 研究成果

已发表/接收论文

- **Zhou H**, Chen K J, Zhang W M, Yu N H. Comments on “Steganography Using Reversible Texture Synthesis”[J]. IEEE Transactions on Image Processing (**TIP**), 2017, 26(4): 16231625. (一区，CCF A)
- **Zhou H**, Chen K J, Zhang W M, Qin C, Yu N H. Feature-Preserving Tensor Voting Model for Mesh Steganalysis[J]. IEEE Transactions on Visualization and Computer Graphics (**TVCG**), 2019. (一区，CCF A)
- **Zhou H**, Chen K J, Zhang W M, Yao Y Z, Yu N H. Distortion Design for Secure Adaptive 3D Mesh Steganography[J]. IEEE Transactions on Multimedia (**TMM**), 2018, 21(6): 13841398. (一区，CCF B)
- **Zhou H**, Chen D D, Liao J, Zhang W M, Chen K J, Dong X Y, Liu K L, Hua G, Yu N H. LG-GAN: Label Guided Adversarial Network for Flexible Targeted Attack of Point Cloudbased Deep Networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**). 2020. (CCF A)
- **Zhou H**, Chen K J, Zhang W M, Fang H, Zhou W, Yu N H. DUPNet: Denoiser and Upsampler Network for 3D Adversarial Point Clouds Defense[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision (**ICCV**). 2019: 19611970. (CCF A)
- **Zhou H**, Chen K J, Zhang W M, Qian Z X, Yu N H. Targeted Attack and Security Enhancement on Texture Synthesis Based Steganography[J]. Journal of Visual Communication and Image Representation (**JVCIR**), 2018, 54: 100107.

4 研究成果

已申请专利

- 3D网格模型隐写方法，申请号：201910146687.2
- 一种基于 PDF 文件的图文相关鲁棒隐写方法及系统，申请号：201910129282.8

待发表论文

- **Zhou H**, Zhang W M, Chen K J, Li W X, Yu N H. ThreeDimensional Mesh Steganography and Steganalysis: A Review[J]. IEEE Transactions on Visualization and Computer Graphics (**TVCG**), 2020. （一区，CCF A）

4 研究成果

已获得荣誉

- 中科院院长奖优秀奖，2020年。
- 中国互联网发展基金会网络安全专项基金网络安全奖学金，2018年。
- 博士生国家奖学金，2019年。
- IJCAI——阿里巴巴天池对抗攻防竞赛防御赛道第四名，获奖人：卞寰宇，周航，周文柏，2019年。
- 中国科学技术大学优秀毕业生，2020年。
- 中电仪器奖学金，2016年。

谢谢各位老师和同学