2018-2019春季
信息隐藏课程
第3讲 隐写分布特性保持

中国科学院信息工程研究所
INSITUTE OF INFORMATION ENGNEERING CAS

SKLOIS
信息安全国家重点实验室

赵险峰

中国科学院信息工程研究所
信息安全国家重点实验室

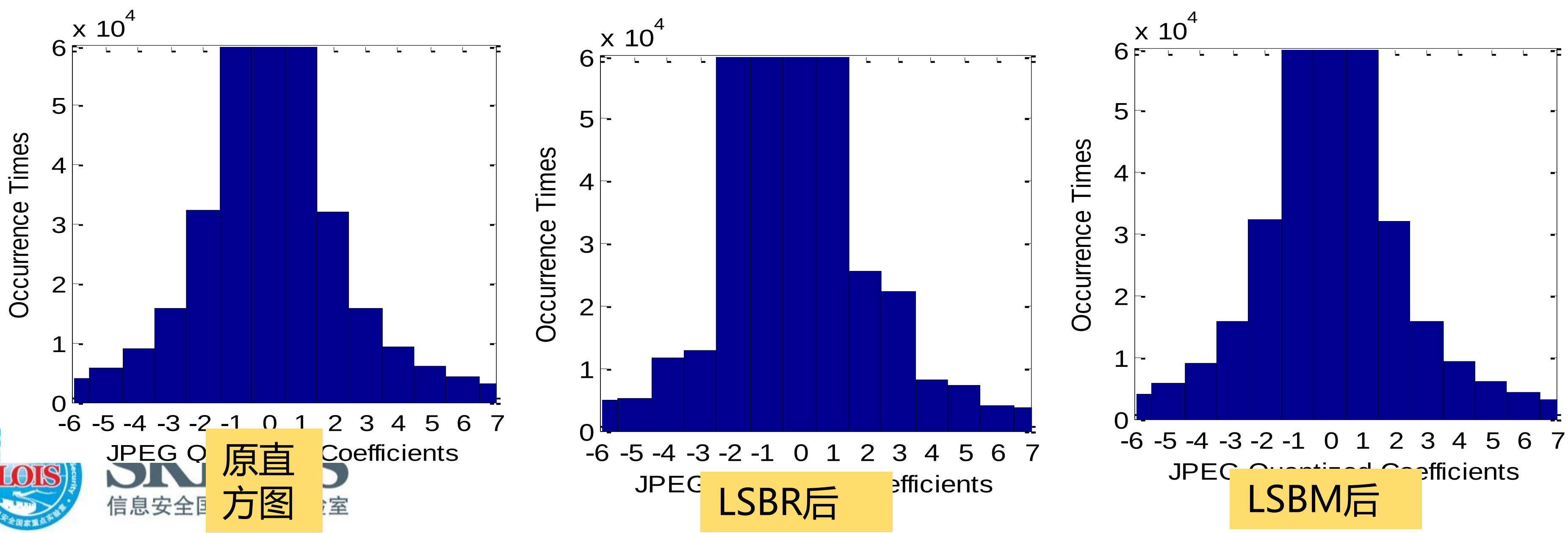
2018年10月

纲要

1. 基本概念
2. 基本分布特性分析
3. 基于分布恢复的统计保持
4. 基于模型的统计保持
5. 基于修改方式的统计保持（下次课）
6. 文献阅读推荐

1-1 上一讲回顾

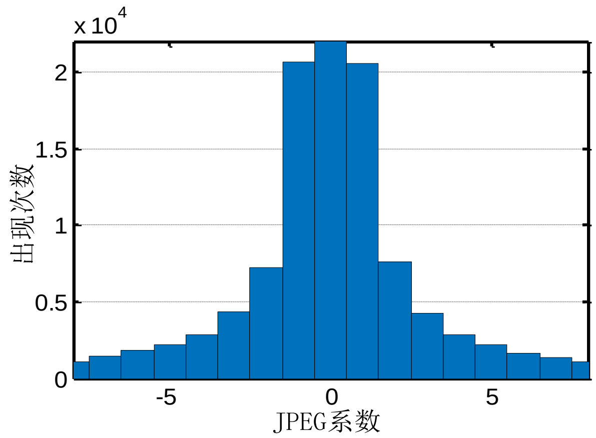
- **上一讲给出了的基本嵌入方法可以构成初级隐写方案，但它们基本并没有专门考虑保持载体统计特征**
  - LSB替换（LSBR）奇数只减、偶数只增，值对分布接近
  - LSB匹配（LSBM）嵌入可以克服“值对接近”现象，但是没有考虑对更多统计特征的保持
  - **本讲介绍的算法不止于克服值对接近现象**

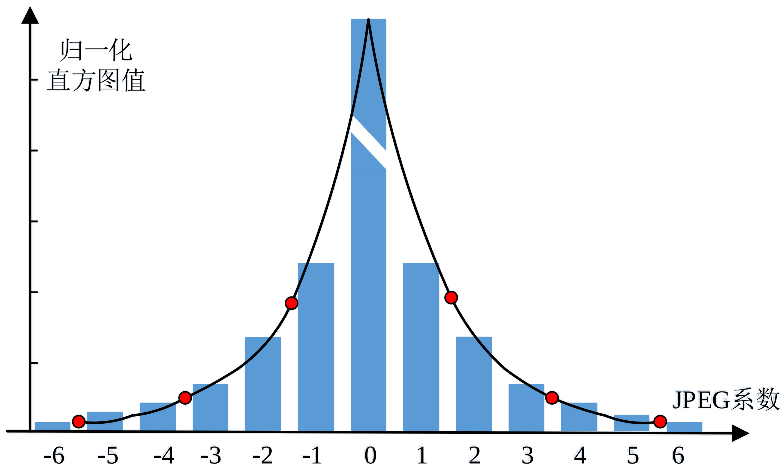


1-2 统计保持概念

- 统计特征保持是指，在隐写中尽可能维持载体的统计分布，而并不是说完全保持原有分布
- 对统计特征的保持难度非常大，存在各阶特征交叉影响
- 由于统计特征数量庞大，当前的主要统计保持方法之一是保护载体的分布，实际这能够保持或有利于保持载体的其他统计特性
- 保护一阶以上载体分布的难度非常大，尚没有出现完全有效的算法，本讲仅给出保护一阶分布的典型方法

2-1 基本分布特性分析——JPEG系数





- 从统计上看，JPEG图像量化系数符合拉普拉斯分布（Lalatian Distribution）或广义柯西分布（Generalized Cauchy Distribution）
  - 对称性。以0值为中心达到最大值，两侧分布对称
  - 单侧单调性。以0值为中心达到最大值，两侧单调下降
  - 大小值分布的比例性。小值较多，但是大致也有一定的比例，表现在分布波形“胖瘦”有度。以后将看到，一些隐写算法会使得波形出现向0值方向的“收缩”
- 一些基本的隐写分析方法主要基于以上特性识别隐写后的JPEG图像，因此，JPEG隐写方法必须满足这些特性的基本约束

2-2 χ^2分析（针对识别LSBR）

- **上讲介绍LSBR时，指出这类隐写造成在奇数值样点上只减，在偶数值样点上反之，使得数值相邻样点的分布密度接近，该特性可基于 χ^2 统计量表征和识别**
- 设 $h(2i)$ 表示载体样点在 $2i$ 处的直方图值（这里限定 $i>0$，$i<0$ 的情况类似），$h^*(2i)$ 为修改后的值，则根据上小节的描述，不失一般性，由于 $h(2i) > h(2i+1)$，则在LSB隐写后，更多的 $2i$ 变为了$2i+1$，因此

$|h(2i)- h(2i+1)| \geq |h^*(2i)- h^*(2i+1)|$ **(对值分布更接近)**

- **如何从统计量上刻画这一特性（异常特征提取）？**

2-3 χ^2分析（值对现象的统计量特征刻画）

- **引入 χ^2 统计特征。** N(0,1)高斯分布变量平方和的分布。记

$$y^*(i)=\frac{h^*(2i)+h^*(2i+1)}{2},\quad y(i)=h^*(2i)$$

- **由于** $h^*(2i)+h^*(2i+1)=h(2i)+h(2i+1)$ **（值对不外流）**，可以通过衡量固定值$y^*(i)$ 与 $y(i)$的距离和进行分析

$$t=\sum_{i=1}^{d-1}\frac{\left(y(i)-y^*(i)\right)^2}{y^*(i)}=\sum_{i=1}^{d-1}\left(\frac{y(i)-y^*(i)}{\sqrt{y^*(i)}}\right)^2=\sum_{i=0}^{d-1}\frac{\left(h^*(2i)-h^*(2i+1)\right)^2}{2\left(h^*(2i)+h^*(2i+1)\right)}$$

  - 0和1值样点经常不用，所以以上 i 从1开始：$i=1,2,\cdots,127$
  - $\left(y(i)-y^*(i)\right)/\sqrt{y^*(i)}$ 的构造考虑了使得平方括弧内的值更~N(0,1)
  - 因此，可认为$t\sim\chi^2(d-1)$，即 t 满足自由度为 $v=d-1$ 的 χ^2 分布

$$f(t)=\begin{cases}\dfrac{t^{(v-2)/2}\mathrm{e}^{-t/2}}{2^{v/2}\Gamma(v/2)}, & t>0\\ 0, & t\le 0\end{cases},\quad 其中，\Gamma(x)=\int_0^{+\infty}u^{x-1}\mathrm{e}^{-u}d_u$$

2-4 χ^2分析（统计量特征识别方法）

- **假设检验。** 由于 t 越小越表示存在隐写，可以设计一个阈值 γ，按照假设检验进行判决。这样漏检率、虚警率为

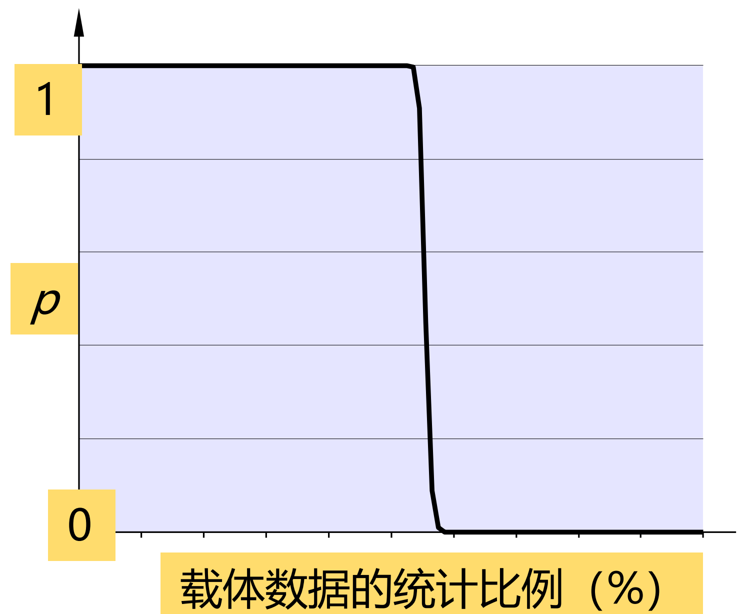
$$P_{\mathrm{MD}}(\gamma)=\int_{\gamma}^{+\infty} f(t)d_t \qquad P_{\mathrm{FA}}(\gamma)=\int_{0}^{\gamma} f(t)d_t$$

$$\text{正确率}=\mathbf{1}-\frac{\text{漏检率}+\text{虚警率}}{\mathbf{2}}=\frac{\text{真阳性率}+\text{真阴性率}}{\mathbf{2}}$$

- 由于对隐写样本 t 的值非常小，因此，可以简单地用以下统计量完成检测：

$$p=\int_{T}^{+\infty} f(t)d_t=1-\int_{0}^{T} f(t)d_t$$

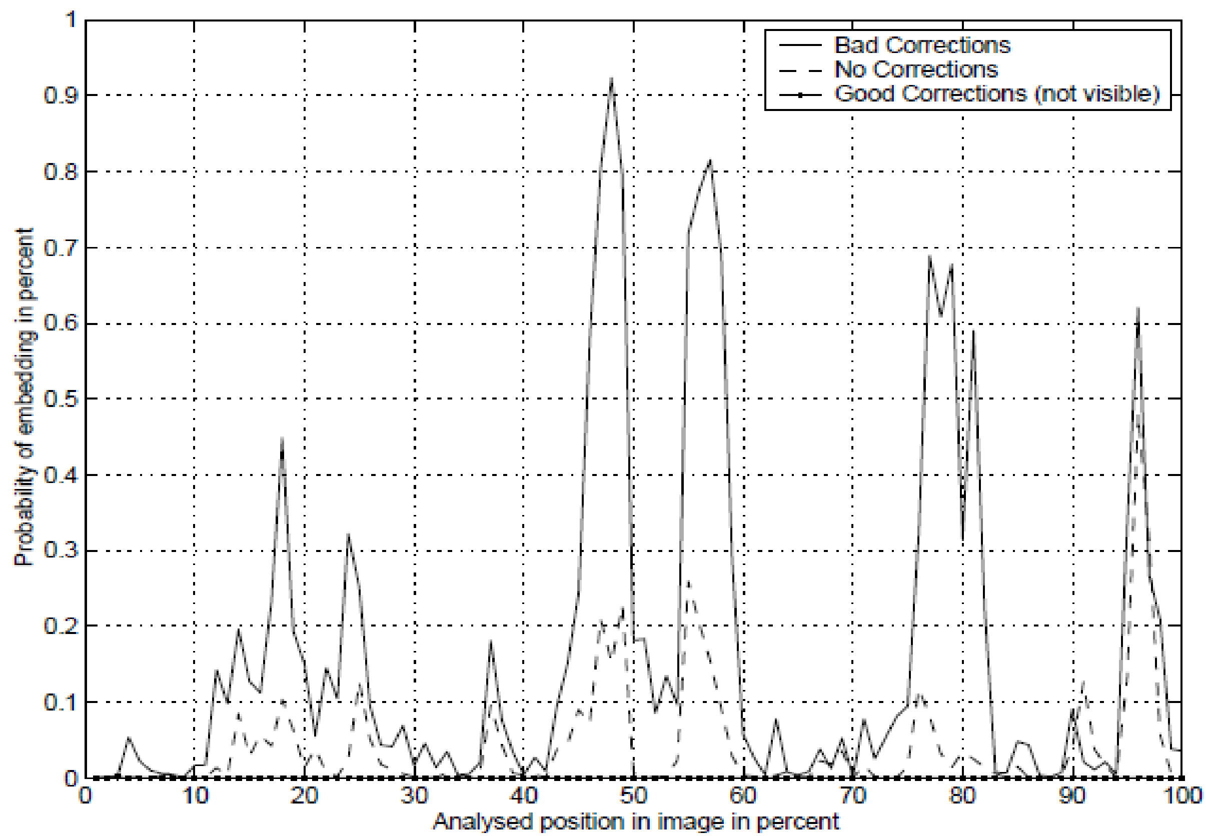
- T 表示具体的 t 值；如果 p 接近1，则认为存在隐写



SKLOIS
信息安全国家重点实验室

2-5 χ^2分析（移动窗口法）

- 实验表明， χ^2分析方法**对不连续的LSBR分析性能下降也快**
- 因此，N. Provos等人提出了**移动窗口 χ^2 分析方法**，对随机选择位置的LSBR，可能锁定一个密集嵌入区进行分析，得到总体判决结果



SKLOIS
信息安全国家重点实验室

3-1 基于分布恢复的统计保持（OutGuess密码方案）

- OutGuess是一款JPEG图像隐写软件，采用了非常典型的一阶统计特征保持方法：在LSB嵌入之后，利用调整非嵌入区的LSB值，修正改变的直方图（一阶分布特征）
- 嵌入位置随机跳跃确定，跳跃距离由伪随机数控制
- 密钥与密码方案
  - OutGuess采用流密码RC4加密待隐藏的消息，密钥（密钥流发生器状态，或称种子）由消息收发双方共享
  - 由另外的嵌入位置选择种子控制，将RC4产生的密钥流也作为伪随机数发生器（Pseudo-Random Number Generator，PRNG）使用，每一段输出作为位置的嵌入跳跃距离（详见下面的讨论）
  - 由消息长度与位置选择种子组成的状态信息也作为消息的一部分
    - 状态信息的嵌入位置由共享密钥驱动RC4密钥流发生器生成

3-2 OutGuess嵌入位置的随机确定

- OutGuess体现了一定的自适应处理思想，它优选32个PRNG种子中的一个，在嵌入前即时产生一个最优的（所需修改量最少的）位置序列
- 设可嵌入的位置依次为 $b_i, i = 1,2,\cdots,m$， m 为需要嵌入的比特数量，在每一个种子下，位置选择方法可以表示为

$$b_0 = 0$$
$$b_i = b_{i-1} + R_i(x)$$

- 其中， $R_i(x)$ 表示在 $[1,x]$ 的区间范围内参照PRNG的输出选择一个跳跃位置偏移，并且每嵌入8比特重新计算区间长度

$$\text{区间长度} \approx \frac{2 \times \text{未使用的载体可嵌入样点数}}{\text{未嵌入的消息长度}}$$

- 以上乘以系数2的原因是，有0.5的概率选择的位置偏移不超过 $x/2$

SKLOIS
信息安全国家重点实验室

3-3 OutGuess位置选择的优化

- **总体需要修改的样点数量取决于种子、消息与载体，在后两者确定的情况下，仅仅取决于种子；OutGuess通过采用32个密钥流种子进行嵌入尝试，选择引发最小修改次数的种子作为状态信息**
- **通过对修改次数方差的估计，可以发现以上优选能够起到一定的作用**
  - 设 p 表示一个样点被修改概率，在 n 个可嵌入样点中隐藏同等数量比特消息所需修改次数为 k，则这个事件满足二项分布 $p_k^{(n)}=\binom{n}{k}p^k(1-p)^{n-k}$。由概率论得知，若 $n=1$，k 满足0-1分布，有

$$E(k)=0\cdot(1-p)+1\cdot p=p，E(k^2)=0^2\cdot(1-p)+1^2\cdot p=p$$

$$D(k)=E(k^2)-[E(k)]^2=p-p^2=p(1-p)$$

  - 若 $n>1$，则 k 满足二项分布，有$E(k)=E(\sum_{i=1}^{n}k_i)=\sum_{i=1}^{n}E(k_i)=np$，$D(k)=\sigma^2=D(\sum_{i=1}^{n}k_i)=\sum_{i=1}^{n}D(k_i)=np(1-p)$
  - 令$q=1-p$，有 $\sigma=\sqrt{npq}$，在理想随机情况下有 $q=p=0.5$，$\sigma=p\sqrt{n}$，实际情况会有偏差。相关文献的实验表明，当 $n=14832$，推算和实测的 σ 分别为60.893与53.123

3-4 OutGuess设计者的其他设想

- OutGuess的设计者在文献中还提出，可以**基于纠错码减小修改次数**，尤其是进行两次嵌入，第一次嵌入秘密信息，第二次是一个可公开信息，目的是在隐写事实被发现后，通过提取第二次嵌入的消息对抗质询，实现所谓“**合理的可否认性**”（Plausible Deniability）
- 但是这些只是一些初步设想，其中，纠错码的使用目的是，使得二次嵌入不改动第一次嵌入的部分样点而仍然有效
- 以上在OutGuess软件中并未采用

3-5 OutGuess统计保持原理

- 若直接采用以上方法隐写，可抵御 χ^2 分析，但滑动窗口分析法仍然可以发现隐写的存在。因此，需要进一步修正样点分布
- 设 α 表示量化DCT系数所承载消息比特数占可嵌入系数数量的比例，则隐写前后相邻值对的直方图值 f 与 $\bar{f}$ 分别变为

$$f^* = f - \frac{\alpha}{2}f + \frac{\alpha}{2}\bar{f} = f - \frac{\alpha}{2}(f - \bar{f}), \quad \bar{f}^* = \bar{f} - \frac{\alpha}{2}\bar{f} + \frac{\alpha}{2}f = \bar{f} + \frac{\alpha}{2}(f - \bar{f})$$

  - 上式中，假设 $f \geq \bar{f}$，分别对应邻值系数中绝对值较小与较大值的直方图，系数1/2是由于一个可嵌入样点的修改可能是0.5
  - 中间表达形式中的第2项表示从本值中“流出”到邻值的数量，第3项表示从邻值“流入”的数量
  - 为了利用较大值未承载信息的区域进行直方图复原，要求预留区样点数要大于 f 的变化数量为$(1-\alpha)\bar{f} \geq \frac{\alpha}{2}(f - \bar{f})$，即 $\alpha \leq \frac{2\bar{f}}{f+\bar{f}}$
  - 说明，要完全修复分布，α 有一定限制，虽然一般安全的嵌入率满足此要求，但是由于载体的丰富性不能保证处处满足

3-6 OutGuess统计保持算法策略

- 针对一对邻值上的**（总体）**分布，算法最终需要通过修改邻值进行修正，但是它并不急着立即这么做，而是对各个值上需要的修改次数进行记录，允许暂时不修改一定的次数，目的是希望等待值对上修改需求的相互抵消；
- 只有需要修改的次数超过以上设置的次数，才调用 $exchDCT$ 函数基于前面考察过的区域通过修改邻值进行修正；
- 但在修正失败的情况下，继续增加记录的修改次数；逐个系数考察完毕后，最后对记录的需要修改次数再进行一轮处理，不确保实现完全的修正

3-7 OutGuess统计保持算法（Part1）

1. $N \leftarrow DCTFreqTable(Original\ Image)$；//计算量化DCT系数的直方图

2. $k \leftarrow Number\ of\ Coefficients\ for\ Embedding$；//获得可嵌入系数样点数

3. $\beta \leftarrow$ 按照经验值设置上限；　//设置允许多大比例不立即修正，而是等待抵消

4. for $v = DCT_{min}$ to DCT_{max} do　//从系数的小值到大值循环

5. 　$N_{err}[v] \leftarrow 0$；　　　　//先设每个DCT值需要被修正的次数为零

6. 　$N^*[v] = \beta N[v]$；　　　　//记录每个DCT值对应的样点暂不修正数量

7. endfor

8. for $i = 1$ to k do　　　　//按系数的位置逐一循环

　// 对未修改（隐写或者修正引起的修改）位置不处理，返回

9. 　if $DCT(i)$ unmodified then

10. 　　Continue；

11. 　endif

　//以下处理修改过的位置

　//对本值⊕1后得到邻值

12. 　$AdjDCT \leftarrow DCT(\mathrm{i}) \oplus 1$；　　　//修改方式对任何系数均是LSB

SKLOIS
信息安全国家重点实验室

3-8 OutGuess统计保持算法（Part2）

if $N_{err}[AdjDCT]$ then //如果记录显示值对邻值还需要$N_{err}[AdjDCT]$次修正

  $N_{err}[AdjDCT]$减1；//则所需修正次数记录减1

  Continue；   //邻值正好也需要修正，则2个值的修正抵消

endif;

//以下处理有修改但没有相互抵消的情况

//如果$DCT(i)$对应值上需要的修正次数<该值上暂时允许不修改的次数，
//则可以继续记录（就是不急着修正，等待着以上抵消出现）

if $N_{err}[DCT(i)] < N^{*}[DCT(i)]$ then

  $N_{err}[DCT(i)]$加1； //$DCT(i)$对应的值上需要的修正次数加1

            //这类需要修改的次数主要由之前的嵌入引起

  Continue；

endif;

3-9 OutGuess统计保持算法（Part3）

// 如果超过缓存不修正的比例上限，通过将空闲区邻值改为本值解决，
// 并且**通过修改前面（$j < i$）一个考察过的本值样点（无修改）完成**

if $exchDCT(i, DCT(i))$ fails then // $exchDCT$先执行，顺利的话修正当前修改
//以下仅仅在$exchDCT$失败时执行，此时只能继续增加记录的未修改比例
  $N_{err}[DCT(i)]$加1；
  Continue；
endif；
endfor;
//遗留缓存中的全部用本值非修改位置上的样点做修正
for $v = DCT_{min}$ to DCT_{max} do
  while $N_{err}[v] \neq 0$ do
    $N_{err}[v]$减1；
    $exchDCT(k, v)$； //修改考察范围是整个系数集，k表示搜索全部范围
  endw
endfor

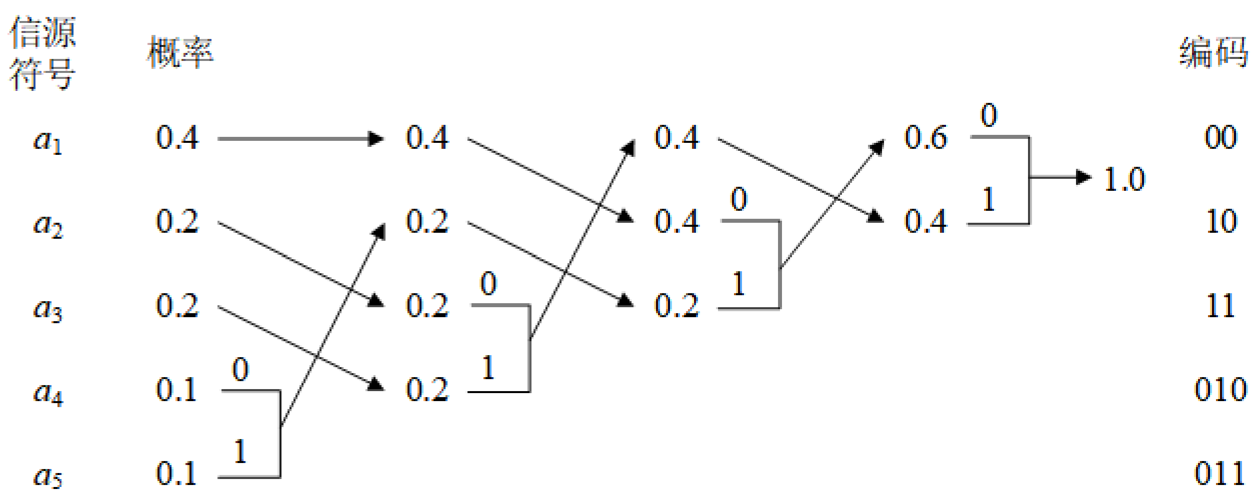
4-1 基于模型的统计保持（动机）

- 以上基于分布恢复的方法显著降低了嵌入效率，实际也更严重影响了二阶及以上阶统计特征。在保持一阶载体分布方面，P. Salle提出了基于模型（Model-based，MB）的隐写，不会引起嵌入效率下降的问题，甚至还有提高
- 回顾：**嵌入效率（Embedding Efficiency）。嵌入效率 e 的意义是，平均每修改一个位置所能传输的隐蔽消息信息量，若信息量用比特表示，则计算公式为**

$$e = \frac{\text{平均每载体样点承载的信息比特}}{\text{平均每载体样点被修改量}} \; bits/\text{次修改}$$

4-2 Huffman编码回顾

- 在变长编码中，若将各码字长度按照所对应信源符号出现概率的大小逆序排列，则码字的平均长度最小。即短码字对应出现概率大的符号，反之亦然
- 基于上述原理，Huffman编码可以用以下步骤描述：
  - 按照递减顺序排列**信源符号**的出现概率
  - 将最小的两个概率相加得到一个新的概率和新的递减列表，每次重复这一过程，直到列表中只剩下概率1.0为止
  - 对每次的组合路径，进行组合路径（二叉树）的描绘， 将概率较大的路径标注为1，将较小的标注为0，或者反之
  - 将以上路径从开始（最低位）直到最终1.0上的标注记录下来，即Huffman码
- 从以上Huffman编码的描述看，**信源符号出现的概率**是**编解码的配置参数**



SKLOIS
信息安全国家重点实验室

4-3 MB隐写的统计保持原理

- MB隐写的框架适合一般的嵌入域
  - 设载体样点为 X，其中不受嵌入影响的信息为 X_α，受影响的信息为 X_β(非1即0)。显然，分布$P(X)$ 主要受 X_α 影响，**收发双方都可用 X_α 通过拟合分布曲线**，得到近似于$P(X)$ 的分布$\hat{P}(X)$。将 X_β修改为X'_β可以认为是对传输1还是0的一种表达信息，因此可以认为 X_β、 X'_β是0或1。
  - 一般情况下，$\hat{P}(X)$ 反映了X 分布曲线的基本轮廓。如果通过修改 $\boldsymbol{x}_\beta$ 为 $\boldsymbol{x}'_\beta$ 嵌入加密消息，发送者希望 $X' = X_\alpha || X'_\beta$ 满足约束（|| 表示连接）：$P(X') \approx \hat{P}(X)$
  - 由于$P(X') = P\left(X'_\beta | X_\alpha\right) P(X_\alpha)$，$P(X_\alpha)$嵌入前后不变，因此约束实际为：

$$P\left(X'_\beta | X_\alpha\right) = \frac{P(X')}{P(X_\alpha)} = \frac{P\left(X_\alpha || X'_\beta\right)}{P(X_\alpha||0) + P(X_\alpha||1)} \approx \hat{P}\left(X'_\beta | X_\alpha\right) = \frac{\hat{P}\left(X_\alpha || X'_\beta\right)}{\hat{P}(X_\alpha||0) + \hat{P}(X_\alpha||1)}$$

  - **所谓基于模型的隐写，就是要求隐写满足以上统计模型约束**
  - 如果基本嵌入为LSBR，以上X_β可以认为是LSB， X_α是不变的其他位平面。J. Fridrich的教材采用这种简化的描述，除了AC系数，也不使用0与1作为嵌入域，这样，绝对值最小的值对是2与3以及−2与−3

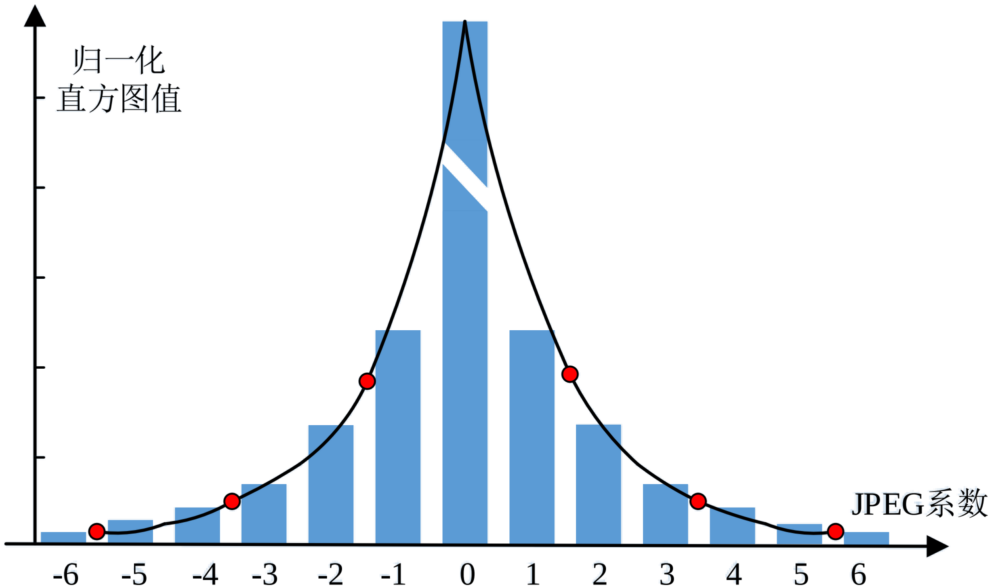
4-4 MB隐写的基本嵌入方法说明

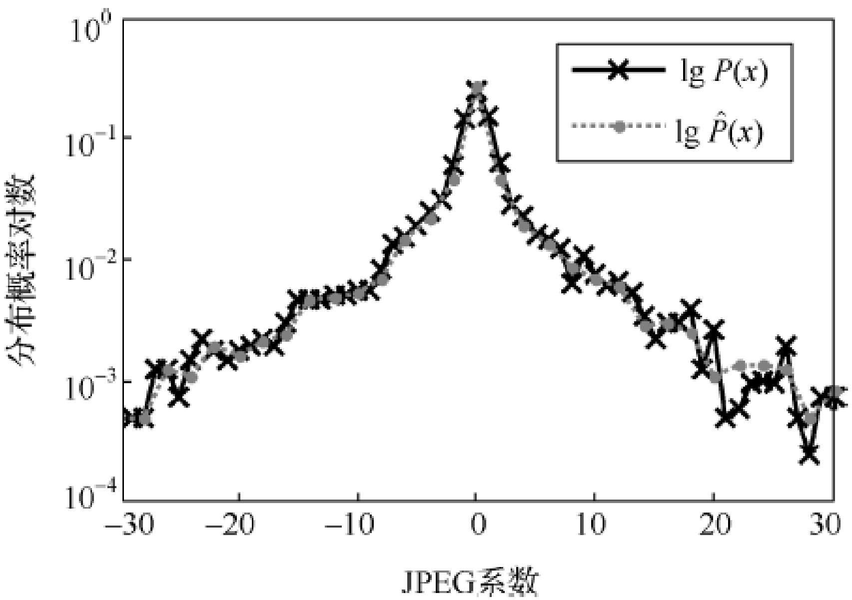
- **原论文中，MB隐写采用了“奇小偶大值对LSBR”的基本修改方法（见前一讲的基本嵌入方法部分），在需要修改时，对奇数绝对值加1，对偶数绝对值减1； MB隐写不使用DC系数与AC系数的0值；这样，绝对值最小的值对是1与2以及−1与−2**
- 如果采用奇小偶大值对LSBR，虽然也改动了次LSB位平面，但是以上约束还是可以成立的，其中，X_α可以认为是一个值对的标识信息，由于值对是封闭对流，值对上的分布不变，这样 $\hat{P}(X)$ 可以通过 X_α估计，$\hat{P}(X'_\beta|X_\alpha)$ 可以由发送者与接收者通过 $\hat{P}(X)$ 计算得到。例如，在奇小偶大值对LSBR基本嵌入下，1与2是一个值对，有约束：

$$P(0|\text{样点属于1或2}) \approx \frac{\hat{P}(\text{样点属于1或2} \,||\, 0)}{\hat{P}(\text{样点属于1或2} \,||\, 0)+\hat{P}(\text{样点属于1或2} \,||\, 1)}$$

$$P(1|\text{样点属于1或2}) \approx \frac{\hat{P}(\text{样点属于1或2} \,||\, 1)}{\hat{P}(\text{样点属于1或2} \,||\, 0)+\hat{P}(\text{样点属于1或2} \,||\, 1)}$$

4–5 $\hat{P}(X)$ 的估计思路





粗点已知（部分点），拟合后可以得到全部点

参数估计，分布曲线拟合

4-6 $\hat{P}(X)$ 的估计方法

- 对每个频率的分块DCT系数均要估计$\hat{P}(X)$，嵌入的消息首先需按DCT系数的类型做一次分割。$\hat{P}(X)$ 的估计采用广义柯西分布参数模型：

$$h(x) = \frac{p-1}{2s}\left(1+\frac{|x|}{s}\right)^{-p}，\ s, p\text{是待估计的参数}$$

- 可认为 X_α 的频次是 $X_\alpha||0$ 与 $X_\alpha||1$ 频次的和。记 $X_\alpha||0$ 为$2i$ 并记 $X_\alpha||1$ 为 $2i+1$（这里设 $i<0$， $i>0$ 的情况类似），则 $h(2i)+h(2i+1)$ 的值可通过统计 X_α 的频次得到，这个数值的一半大约是 $(2i+2i+1)/2$的频次，即有

$$h\left(\frac{2i+2i+1}{2}\right)=\frac{h(2i)+h(2i+1)}{2}$$

- 因此，可基于以上得到的离散点（称为低分辨率直方图）拟合得到分布；需要指出，不准备采用的DCT量化系数值（如0）单独统计。按照以上参数模型，通过最大似然估计法得到 $h(x)$ 作为 $\hat{P}(X)$。基于$h(x)$，有：

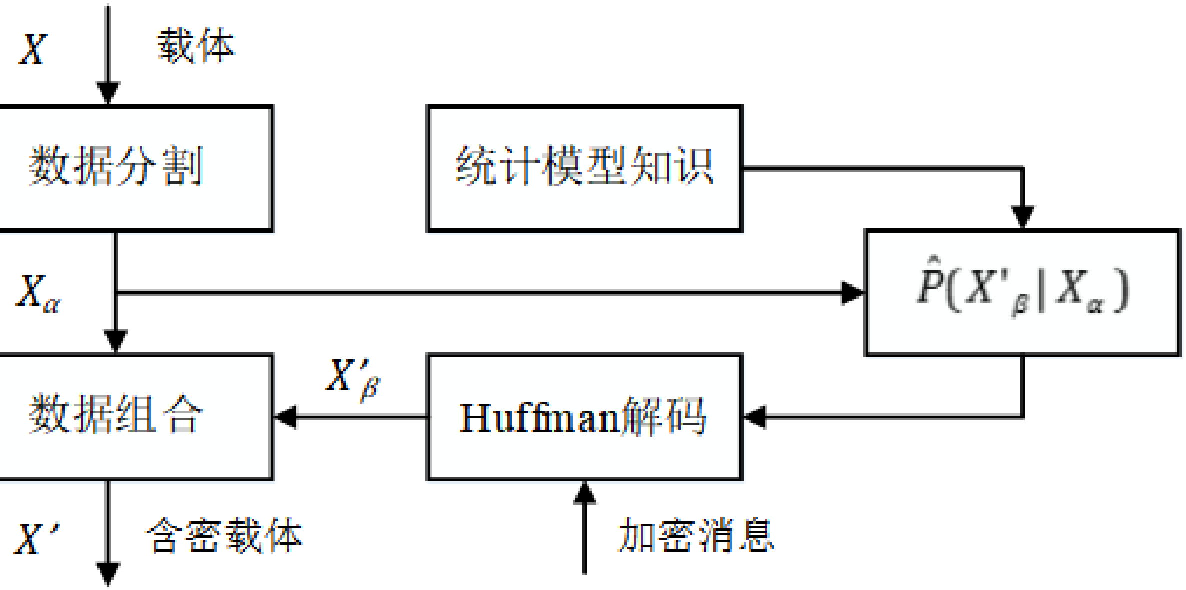
$$\hat{P}\left(X'_\beta=0|X_\alpha=2i\right)=\frac{h(2i)}{h(2i)+h(2i+1)},\quad \hat{P}\left(X'_\beta=1|X_\alpha=2i\right)=\frac{h(2i+1)}{h(2i)+h(2i+1)}$$

4-7 基于MB的隐写与统计保持

- 由于隐写的发送者能够根据 X_α 估计$\hat{P}(X)$并进一步计算$\hat{P}\left(X'_\beta|X_\alpha\right)$，因此，可以结合利用Huffman解码的特性，基于以下JPEG图像隐写嵌入加密消息并满足以上模型约束

1. 确定嵌入区域和位置置乱。对JPEG载体图像，进行无损压缩的解压，得到分块DCT量化系数；在每个分块DCT量化系数中，选择AC系数作为嵌入域，后者中，数值为0的系数也将不用于嵌入，基本的嵌入方法是奇小偶大值对LSBR；在嵌入前，需要对全部可嵌入系数进行位置上的置乱
2. 分布估计。对**不同频率（Mode）**的分块AC系数，按照前述的方法基于 $\boldsymbol{x}_\alpha$ 估计$\hat{P}(X)$
3. 系数分组。用于嵌入的系数有两次分组，第一次基于**不同的频率**分为大组，第二次在每个大组中基于**不同** $\boldsymbol{x}_\alpha$ 的值分为小组，消息将按照最后得到的小组分片嵌入。因此，以下过程（4）与（5）针对每个 $\boldsymbol{x}_\alpha$值执行一次
4. 概率计算。根据$\hat{P}(X)$ 与前述方法计算 $\hat{P}\left(X'_\beta|X_\alpha=\boldsymbol{x}_\alpha\right)$
5. 密文嵌入。将 $\hat{P}\left(X'_\beta|X_\alpha=\boldsymbol{x}_\alpha\right)$ 作为构造Huffman解码器的基础概率，构造相应的解码器；将待传输的密文流 X'_β作为一个Huffman压缩流解压到 X_β的位置上，嵌入修改方法是奇小偶大值对LSBR。
6. 反置乱与压缩编码。将以上在置乱域嵌入的DCT量化系数进行位置上的反置乱，在进行无损压缩（熵编码），得到隐文JPEG图像。

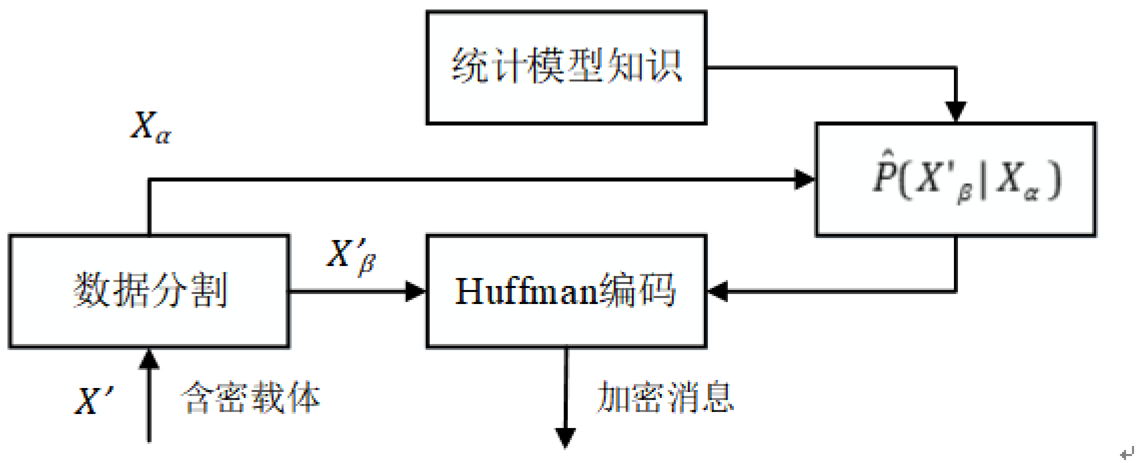
4-8 基于MB的隐写与统计保持（图示）



4-9 MB隐写的消息提取

- 由于隐写的接收者能够根据不变的 X_α 估计 $\hat{P}(X)$ 并进一步计算 $\hat{P}(X'_\beta|X_\alpha)$，因此，可以结合利用Huffman编码的特性，基于以下算法提取以上隐写嵌入的加密消息：

1. 类似地执行以上嵌入方法的（1）至（4）；
2. 密文提取。将 $\hat{P}(X'_\beta|X_\alpha=\boldsymbol{x}_\alpha)$ 作为构造Huffman编码器的基础概率，构造相应的编码器，将 X'_β 作为输入流输入编码器，得到的Huffman编码流作为提取的密文。



4-10 MB算法的嵌入效率

- 由于经过了压缩编解码器的处理，信息流的尺寸发生了改变，因此需要借助信息论的原理分析嵌入效率。
  记 $s = P\left(X'_{\beta} = 0 | X_{\alpha}\right)$，则在一个可嵌入系数上传输的信息量是以下的熵值：
  $$H(s) = -s\log_2 s - (1-s)\log_2(1-s)$$
- 从全部分组可嵌入系数的情况看，LSB嵌入前后为0的概率相等，嵌入前后为1的概率相等，因此，修改的概率（平均每样点上的修改次数，仅统计0到1与1到0两个情况）为$s(1-s)+(1-s)s = 2s(1-s)$，因此，嵌入效率为
  $$e = \frac{-s\log_2 s - (1-s)\log_2(1-s)}{2s(1-s)}$$
- 从数值分析结果看，这个值除了当 $s = 0.5$ 时均大于2；由于MB保持了原始的分布特性，因此 s 更可能与0.5有更大的偏移，这个是区别于将LSB直接替换为密文的

5 基于修改方式的统计保持（下次课）

- LSB匹配（LSB Maching，LSBM）嵌入有助于克服出现以上的 χ^2 特征，是基于修改方式的统计保持的简单例子
  - 在LSBM嵌入中，也是用最后的LSB承载秘密消息，但是，当需要修改LSB的值时，LSBM嵌入时通过对样点值做随机的加减1
- F3、F4 （下次课与F5一并介绍）通过选择修改方式，保护了JPEG系数的分布特性

6 文献阅读推荐

- [1] 教材第3章
- [2] N. Proves. Defending against statistical steganalysis. The 10th USENIX Security Symposium, Washington, DC, USA, August 2001, pp.323-335
  - 描述OutGuess隐写及其基于统计恢复修正的统计保持方法
- [3] P. Sallee. Model-based steganography. In Proc. IWDW’03, LNCS 2939, pp. 154–167, Springer-Verlag, 2004.
  - 描述基于模型（MB）的隐写及其统计保持方法

谢谢！

中国科学院信息工程研究所
INSITUTE OF INFORMATION ENGNEERING CAS

SKLOIS
信息安全国家重点实验室