

基于逆问题扰动的脆弱数字水印认证

赵险峰,汪为农,陈克非

(上海交通大学计算机科学与工程系,上海 200030)

摘要: 为改进基于脆弱数字水印的多媒体数据认证的性能,利用逆问题的扰动现象提出了一种新的脆弱水印体制.在新的方法中,数据完整性或签名的验证并不依赖于水印的提取,而通过反向求解植入方程完成.由于扰动现象的存在,在数据被篡改的情况下,反向求得的数据值将产生猛烈的增长,并且,扰动值反映了篡改的程度,扰动区域正好描述了篡改的轮廓.在这一机制下,脆弱水印还可以引入自适应植入算法,在被保护数据的每一码字上植入水印,与常用的基于分块的算法相比,所提出的算法在提高篡改敏感性和可定位性的同时,还兼顾了感知透明性.

关键词: 脆弱数字水印; 逆问题; 数据认证; 信息隐藏

中图分类号: TP309; TP37 **文献标识码:** A **文章编号:** 0372-2112 (2002) 12A-2130-04

Fragile Watermarking Authentication Based on the Perturbation of Reverse Problems

ZHAO Xian-feng, WANG Wei-nong, CHEN Ke-fei

(Dept. of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: To enhance the performances of multimedia authentication based on fragile watermarking, a new fragile watermarking scheme, which exploits the perturbation of reverse problems, is proposed. To verify data integrity or authenticate signature, the new approach just solves the embedding equation instead of really extracting the embedded watermark. Because of the existence of the perturbation, the value of the reversed solution increases violently if any tampering happened. The perturbed values not only indicate the level of the tampering, but also directly draw the shapes of the manipulated areas. Exploiting this mechanism, fragile watermarking can also use perceptually adaptive embedding, and hide watermark at every codeword of the protected data. Compared with the mostly used block-based methods, the proposed algorithm achieves the better performances of tampering sensitivity and alteration localizability, and in the meantime enhances the perceptual transparency.

Key words: fragile digital watermarking; reverse problems; authentication; information hiding

1 引言

数字签名是数据认证常用的方法,然而,脆弱数字水印 (fragile digital watermarking) 在某些场合却更适用^[1]. 数字签名产生的附加信息必须在特定的协议框架下与原始数据保持对应关系,并且仅依靠签名难以识别篡改位置. 与数字签名不同,脆弱水印在不明显影响发布数据感知质量的情况下,将水印隐蔽地植入原始数据,并通过分析提取水印的变化识别篡改位置. 当前,这项技术主要用于多媒体数据的认证,由于它对认证协议的简化和对篡改位置的分析能力,正在成为数据认证的重要手段之一^[1~4].

在现有技术路线下,脆弱水印主要的性能要求在一定程度上呈现出相互制约的一面. 篡改敏感性要求算法能检测数据尽可能小的改动;可定位性要求算法能识别尽可能小的篡改位置;感知透明性要求水印的植入对人的感觉没有影响或影响在可接受的范围内;盲性 (blindness) 要求水印检测不依赖于原始数据;安全性要求算法必须建立在密钥的基础上. 为实

现上述性能,当前大多数方法将原始数据分块,逐块均匀地植入水印^[2~4]. 然而,篡改敏感性和可定位性的提高依赖于小的分块,而感知透明性和安全性的提高又依赖于大的分块,使得分块尺寸难以把握^[2];另外,盲性的要求使这些方法不能引入自适应植入算法^[5],并且在篡改敏感性和可定位性方面有所损失^[3,4].

为进一步探讨满足脆弱水印性能要求的合适途径,本文基于逆问题的扰动现象^[6],对脆弱水印算法体制进行了新的分析和设计.

2 存在的问题和反向扰动的引入

为了获得提高性能的基本途径,这里先给出脆弱水印算法的一般形式. 设 x 表示原始数据或其某一变换域, w 表示水印信息, k 表示密钥,则脆弱水印算法可以一般化地表示为

$$\begin{cases} \text{embedding: } x'_i = \text{emb}(x_i, \text{ciph}(w_i, k_i)) \\ \text{extraction: } w'_i = \text{deciph}(\text{extr}(x'_i), k_i), \end{cases} \quad i = 0, 1, \dots, n_b - 1$$

(1)

这里, \mathbf{x} 、 \mathbf{w} 和 \mathbf{k} 分别被划分为 n_b 个子块或子段(当不使用流密码时 \mathbf{k} 不被分块), 随后的植入和验证均以块为单位进行; $\text{ciph}(\cdot, \cdot)$ 是水印的加密算法, $\text{deciph}(\cdot, \cdot)$ 是解密算法; $\text{emb}(\cdot, \cdot)$ 逐块将水印植入原始数据, $\text{extr}(\cdot)$ 逐块提出加密的水印. 显然, 认证结论可以基于提取的水印 \mathbf{w}' 给出. 将式(1)中的抽象算法具体化, 可以概括当前的一些典型算法^[2-4], 然而, 这些算法在技术上还存在一些棘手的问题:

(1) 分块尺寸问题: 可定位性和敏感性需要小的分块, 而感知透明性和安全性需要大的分块尺寸^[2];

(2) 不确定性问题: 在分块尺寸不够大时, 水印的提取不够准确. 在文献[2]中, 由于分块尺寸和感知透明性的限制, Hash 值的长度被迫为 64bits, 大大低于低碰撞率需要的 128bits^[7]. 在文献[3]和[4]中, 算法按分块的统计特性来判断提取水印的比特值, 在发生篡改时, 具有一定的不确定性;

(3) 植入强度分布问题: 在等尺寸分块和均匀植入水印的情况下, 要保证一定的感知质量和水印信息量非常困难; 在盲性的约束下, 当前依靠提取水印的方法不能在检测端重新基于原始数据生成幅调因子^[5], 难以利用自适应植入算法;

(4) 植入域选择问题: 在编码域以外的域中植入水印, 由于通过变换和反变换引入了浮点数运算, 可能会对植入水印造成损害, 而避免这种损害的方法比较复杂^[3,4].

本文即将提出的方法不依赖于原始数据的分块, 而是基于逆问题的扰动现象^[6], 这里先给出基本思路. 设 $h(\cdot)$ 表示自适应植入中生成幅调因子 \mathbf{a} 的算法, 则水印植入可以表示为

$$\mathbf{x}' = \mathbf{x} + h(\mathbf{x}) \otimes \text{ciph}(\mathbf{w}, \mathbf{k}) = \mathbf{x} + \mathbf{a} \otimes \mathbf{c} = \mathbf{x} + \mathbf{s} \quad (2)$$

本文称上式为植入方程, 其中, \otimes 表示两个向量的 Hadamard 乘法, 可以定义为 $\mathbf{a} \otimes \mathbf{s} = (a_0 s_0, a_1 s_1, \dots, a_{n-1} s_{n-1})$, \mathbf{c} 表示水印密文, \mathbf{s} 表示其幅调后的形式. 这里可以不对原始数据进行分块, 植入和检测均以码字为单位, 而感知透明性由 $h(\cdot)$ 保证; 检测端不用提取植入的水印, 而是在将 \mathbf{x} 视为未知量的情况下求解式(2), 直接检测和定位篡改. 上述构想是基于如下的事实^[8]: 在适当的设计下, 如果 \mathbf{x}' 被篡改, 求得的解将遭受强烈的扰动, 而扰动的区域正是篡改的区域. 显然, 在提高篡改敏感性和可定位性的同时, 这种方法还在保持盲性的前提下支持了自适应植入. 本文以下将讨论设计这一类算法的主要问题.

3 基于反向扰动的脆弱水印设计

本文仅讨论线性自适应的水印植入, 但得到的一般性结论也适用于非线性植入的情况.

3.1 自适应的水印植入和植入方程的求解

为方便以下的讨论, 这里以矩阵形式^[9]表达线性自适应水印的算法模型(图 1). 对一维情况, 设原始数据为 $x(n)$, $n = 0, 1, \dots, l-1$, 滤波器系数为 $h(n)$, $n = 0, \pm 1, \dots, \pm l_h$. 由于滤波器输出中由卷积引入的扩展部分一般并无实际意义, 因此幅调因子 $\mathbf{a}(n)$ 和原始数据 $\mathbf{x}(n)$ 的尺寸是相同的, 可以表示为

$$\mathbf{a}_{l \times 1} = \mathbf{H}_{l \times 1} \cdot \mathbf{x}_{l \times 1} \quad (3)$$

$$\begin{aligned} \text{其中 } \mathbf{a}_{l \times 1} &= [a(0) a(1) \cdots a(l-1)]^T, \mathbf{x}_{l \times 1} \\ &= [x(0) x(1) \cdots x(l-1)]^T, \end{aligned}$$

$$\mathbf{H}_{l \times l} = \begin{bmatrix} h(0) & \cdots & h(-l_h) \\ \vdots & \ddots & \vdots \\ h(l_h) & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & h(l_h) \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & h(0) \end{bmatrix}$$

$$\triangleq \text{toep}[[h(0) \cdots h(-l_h)], [h(0) \cdots h(l_h)]^T]_{l \times l} \quad (4)$$

以上 $\mathbf{H}_{l \times l}$ 是由 $h(n)$ 生成的 Toeplitz 矩阵, $[\cdot]^T$ 表示矩阵的转置, \triangle 表示“记为”. 这里的扩展方法可以是补零扩展(zero-padding)或对称扩展(symmetrical extension)等, 当使用后者时, 对应于 $\mathbf{H}_{l \times l}$ 外的滤波器系数可以通过折叠加回到 $\mathbf{H}_{l \times l}$. 设水印密文 $\mathbf{c}(n)$ 的向量形式为 $\mathbf{c} = [c(0) \ c(1) \ \cdots \ c(l-1)]$, 为将式(2)中的 Hadamard 乘法用矩阵乘法替换, 定义其对角矩阵形式为 $\hat{\mathbf{C}}_{l \times l} = \text{diag}(\mathbf{c})$, 其中, 当 $i \neq j$ 时, 有 $\hat{c}_{i,j} = 0$, 否则 $\hat{c}_{i,j} = c_i$. 这样, 设 \mathbf{E} 为单位矩阵, 则植入方程可以表示为

$$\begin{aligned} \mathbf{x}'_{l \times 1} &= \hat{\mathbf{C}}_{l \times l} \cdot \mathbf{a}_{l \times 1} + \mathbf{x}_{l \times 1} \\ &= \hat{\mathbf{C}} \cdot (\mathbf{H} \cdot \mathbf{x}) + \mathbf{x} = \mathbf{s} + \mathbf{x} = (\hat{\mathbf{C}} \cdot \mathbf{H} + \mathbf{E}) \cdot \mathbf{x} \quad (5) \end{aligned}$$

上式的计算代价 $O(n^2)$ 一般是可以接受的. 由于式(5)的系数矩阵 $\hat{\mathbf{C}} \cdot \mathbf{H} + \mathbf{E}$ 是非常窄的带状阵, 并且通常是对角占优的, 所以一般是是非奇异的, 即便不是这样, 也可以通过对 \mathbf{H} 的简单设计使之非奇异, 所以本文假设 $(\hat{\mathbf{C}} \cdot \mathbf{H} + \mathbf{E})^{-1}$ 存在. 这样, 植入方程的解可以表达为

$$\mathbf{x}' = (\hat{\mathbf{C}} \cdot \mathbf{H} + \mathbf{E})^{-1} \cdot \mathbf{x}' \quad (6)$$

这里 \mathbf{x}' 可能已经遭受篡改, \mathbf{x}' 可能产生不同程度的扰动.

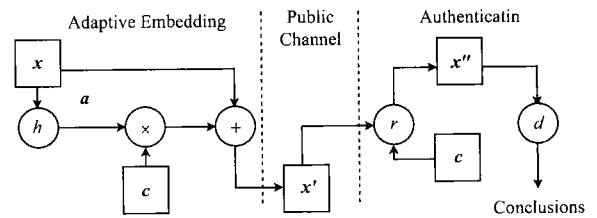


图 1 自适应植入和基于求解的认证 (\mathbf{x} : 原始数据, \mathbf{x}' : 发布版本, \mathbf{x}'' : 反解版本, \mathbf{c} : 水印密文, \mathbf{a} : 幅调因子, h : 滤波, r : 解方程, d : 扰动分析)

在进一步讨论前, 这里先将上述模型转换为二维形式^[9]. 设原始数据为 $x(m, n)$, 其中, $m = 0, 1, \dots, u-1$, $n = 0, 1, \dots, v-1$, 二维滤波器的系数为 $h(m, n)$, 其中, $m = 0, \pm 1, \dots, \pm l_{hu}$, $n = 0, \pm 1, \dots, \pm l_{hv}$. 这样, 式(3)的二维形式是

$$\mathbf{a}_{uv \times 1} = \mathbf{H}_{uv \times uv} \cdot \mathbf{x}_{uv \times 1} \quad (7)$$

其中,

$$\mathbf{a}_{uv \times 1} = [a_0^T \ a_1^T \ \cdots \ a_{u-1}^T]^T, \quad (8)$$

$$\mathbf{a}_i = [a(i, 0), a(i, 1), \dots, a(i, v-1)]$$

$$\mathbf{x}_{uv \times 1} = [\mathbf{x}_0^T \ \mathbf{x}_1^T \ \cdots \ \mathbf{x}_{u-1}^T]^T, \quad (9)$$

$$\mathbf{x}_i = [x(i, 0), x(i, 1), \dots, x(i, v-1)]$$

$$\mathbf{H}_{uv \times uv} = \text{toep}[[H_0 \cdots H_{-l_{hu}}], [H_0 \cdots H_{l_{hv}}]^T]_{u \times u}$$

$$H_i = \text{toep}[[h(i,0)\cdots h(i, -l_{hu})], [h(i,0)\cdots h(i, l_{hv})]^T]_{v \times v} \quad (10)$$

这里, H_i 由 $h(m, n)$ 的第 i 行生成. 与一维的情况类似, 水印密文 $c(m, n)$ 的矩阵形式

$$C_{u \times v} = [c_0 \quad c_1 \quad \cdots \quad c_{u-1}]^T, \quad (11)$$

$$c_i = [c(i,0) \quad c(i,1) \quad \cdots \quad c(i, v-1)]$$

也必须转换为对角形式

$$\hat{C}_{w \times w} = \text{diag}(\text{diag}(c_0) \quad \text{diag}(c_1) \quad \cdots \quad \text{diag}(c_{u-1})) \quad (12)$$

后才能使用. 最后, 式(5)和式(6)均可以表示为其二维形式.

3.2 利用扰动的性质设计篡改的检测和定位

根据逆问题的扰动理论, 通过适当的设计, 在发布版本 x' 遭受篡改的情况下, 可以使植入方程的解 x'' 较原始数据 x 产生较大的偏差. 就前面提出的模型, 求解线性方程组中存在的扰动现象^[8]可以被用于篡改的检测和定位.

植入方程系数矩阵的条件数(condition number)可以用于预估扰动的程度. 对任意方阵 A , 其条件数的定义为 $\text{cond}(A) = \|A^{-1}\| \cdot \|A\|$, $\|\cdot\|$ 表示求矩阵或向量的 Euclidean 范数^[8]. 设 A 代表 $\hat{C} \cdot H + E$, 对植入方程 $A \cdot x' = x$, 设 x' 变化了 $\delta x'$, 则根据线性方程组的扰动理论, 解 x'' 偏离真实解 x 的程度可以通过下式衡量:

$$\|x'' - x\| \leq \|x\| \cdot (\|\delta x'\| / \|x'\|) \cdot \text{cond}(A) \quad (13)$$

当 $\text{cond}(A)$ 足够大时, 求得的解 x'' 一般和原始数据 x 和发布版本 x' 都很不相似, 人的感官一般即可发现这种差异(图 2(d)).

通常植入算法必须将已经植入水印的数据进行量化编码, 量化误差 δq 将引发一个相对轻微的伴随扰动. 为了确保不发生虚警, 这个伴随扰动的上界必须被设为检测阈值. 由植入方程的线性性, 确由篡改引发的扰动可以通过下式衡量:

$$\|x\| \cdot (\|\delta q\| / \|x'\|) \cdot \text{cond}(A) \leq \|x'' - x\| \leq \|x\| \cdot (\|\delta q\| / \|x'\| + \|\delta x'\| / \|x'\|) \cdot \text{cond}(A) \quad (14)$$

由于原始数据 x 和发布数据 x' 非常相似, 一般可以认为有 $\|x\| \approx \|x'\|$ 和 $\|x'' - x\| \approx \|x'' - x'\|$, 则式(14)可以简化为

$$\|\delta q\| \cdot \text{cond}(A) \leq \|x'' - x'\| \leq (\|\delta q\| + \|\delta x'\|) \cdot \text{cond}(A) \quad (15)$$

这样检测算法可以不参照原始数据. 以上基于计算范数的方法显然是不方便的, 但各元素的值可以认为是范数的分量长度, 所以它们由量化误差引发的扰动也是小于总体扰动且有上界的. 我们通过统计发布版本及植入方程的解之间的最大差值来确定这个作为检测阈值的上界 T (图 3(b)中的 $mSlu_Rls$).

为了获得对篡改的定位方法, 还必须讨论扰动的局部性. 由于滤波器的卷积尺寸远远小于原始数据的尺寸, x 中的局部元素仅仅与 x' 中相应位置的元素相关, 但是这种关系是否在 x'' 和 x' 之间也成立? 克莱姆法则(Cramer's rule)^[8]能简要地说明这一问题. 设 $x' \rightarrow A_i$ 是 A 的第 i 列被 x' 置换后的形式, $|A|$ 表示矩阵的行列式, 则有

$$x'' = [|x' \rightarrow A_0|, |x' \rightarrow A_1|, \dots, |x' \rightarrow A_{i-1}|]^T / |A| \quad (16)$$

因为 A 是很窄的带状阵, x' 中被篡改的区域如果不能在 $x' \rightarrow A_i$ 中与对角线相交, 就不能对 x'' 的第 i 个元素产生影响, 据此可以认为, 篡改的区域即为扰动的区域.

4 图像实验和数据分析

实验选用的滤波器兼顾了其一般性和对 $\text{cond}(\hat{C} \cdot H + E)$ 的影响(表 1). 一般情况下, 滤波器的输出反映了局部输入的变化程度^[5], 所设计的滤波器通过计算邻点灰度差的均值来生成幅调因子, 一维情况下, 它为 $[(x_{m-1} - x_m) + (x_{m+1} - x_m)]/2$, 这样, 滤波器系数可以表示为 $h = [1/2 - 1 \quad 1/2]$; 当在两个方向上使用它时, 容易构造一个可分离的二维滤波器, 其系数的矩阵形式为 $H(\gamma) = \gamma h^T \cdot h$, 其中 γ 控制整体上的植入强度.

表 1 10 次实验中相应的条件数(γ : 总的植入强度调节; c : $\text{cond}(\hat{C} \cdot H + E)$, 其中 H 根据式(10)按照 $H(\gamma)$ 得到)

γ	0.25	0.275	0.3	0.325	0.35	0.375	0.4	0.425	0.45	0.475
c	18.9	35.08	530	19554	23438	62209	36084	32002	28655	35647

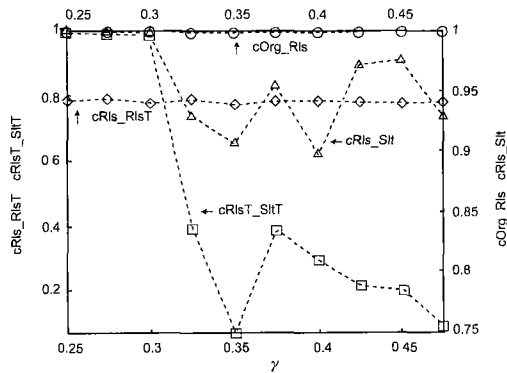
图 2 给出了一组典型的实验图像和基本的实验步骤. 图 2(a)和图 2(b)显示了量化误差引发的轻微扰动; 图 2(d)表示, 在扰动帮助下, 靠肉眼即可判别篡改区域; 虽然图 2(e)说明只有部分扰动值超过了检测阈值, 但这足以说明篡改的发生; 另外, 可以象图 2(f)那样适当降低阈值以获取更详细的篡改信息.



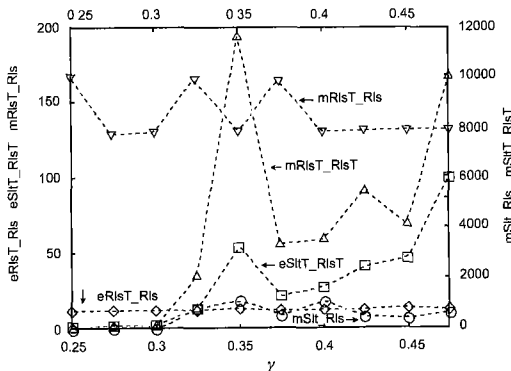
图 2 一组图像实验数据($\gamma = 0.45$; 256 级灰度; $T = 437.7$)

图 3 显示了在 10 次实验中获得的一些有意义的参量值. 在图 3(a)中, $cOrg_Rls$ 非常接近于 1, 说明自适应植入对感知质量的保护作用; $cRls_Slu$ 的值高于 0.9, 说明伴随扰动相

比由篡改引发的扰动要轻微得多; $cRlsT_SlT$ 的迅速下滑表明了在大条件数下,扰动能迅速加大篡改数据和植入方程的解之间的差异。从最大值和均值的角度,图 3(b)进一步说明,在大条件数下,反向求解不仅能加大篡改前后相关参量之间的差异,也能加大反解前后相关参量之间的差异。



(a) 相关系数 ($cOrg_Rls$:原始数据和发布数据之间; $cRls_SlT$:发布数据及其反解版本之间; $cRls_RlsT$:发布数据和被篡改版本之间; $cRlsT_SlT$:被篡改版本及其反解版本之间)



(b) 均值和最大值 ($eRlsT_Rls$:被篡改版本和发布版本差的均值; $eSlT_RlsT$:被篡改版本及其反解版本差的均值; $mRlsT_Rls$:被篡改版本和发布版本差的最大值; $mSlT_Rls$:发布版本及其反解版本差的最大值; $mSlT_RlsT$:被篡改版本及其反解版本差的最大值)

图 3 10 次实验中的一些参量结果 (全为绝对值;当涉及篡改版本或其反解版本时,仅对图 2(c)中方框内的区域进行统计)

5 结论

通过对逆问题扰动的分析,我们发现该现象可以作为设计和提高脆弱数字水印算法体制所依赖的机制。首先,基于反向扰动的方法完全兼容自适应植入算法,使得感知透明性得以充分保证;其次,由于发生扰动的区域正是发生篡改的区域,使得没有必要对原始数据进行任何分块,而直接以码字为定位单位,提高了可定位性;第三,对任何篡改引发的扰动,由于其程度远远大于由量化误差引发的扰动,所以篡改的敏感性和检测的准确性可以得到保证,并且在确认篡改发生的前

提下适当降低检测阈值,可以发现更多的篡改细节;第四,通过反向求解植入方程,可以在不直接提取水印的情况下进行篡改验证,从而在支持自适应植入算法的同时保证了盲性;最后,算法的安全性可以建立在加密的水印上。

虽然本文仅进行了图像实验,但结论显然也适用于音频和视频等其它多媒体数据的认证。

参考文献:

- [1] Lin E T, Delp E J. A review of fragile image watermarks[A]. Proc. of Multimedia and Security Workshop at ACM Multimedia 1999[C]. Orlando, Florida: ACM, 1999.
- [2] Wong P W. A public key watermarking for image verification and authentication[A]. Proc. of IEEE Inter. Conf. on Image Processing[C]. Chicago, Illinois: IEEE, 1998.
- [3] Kundur D, Hatzinakos D. Towards a telltale watermarking technique for tamper-proofing[A]. Proc. of IEEE Inter. Conf. On Image Processing[C]. Chicago, Illinois: IEEE, 1998.
- [4] Lin C Y, Chang S F. Semi-fragile watermarking for authenticating JPEG visual content[A]. SPIE Proc. of Inter. Conf. on Security and Watermarking of Multimedia Contents II[C]. San Jose, California: SPIE, 2000.
- [5] Podilchuk C I, Zeng W. Image-adaptive watermarking using visual models[J]. IEEE Journal on Selected Areas in Communications, 1998, 16(4): 525 - 539.
- [6] Kirsch, A. An Introduction to the Mathematical Theory of Inverse Problems[M]. Berlin: Springer-Verlag, 1996.
- [7] Menezes J, van Oorschot P C, Vanstone S A. Handbook of Applied Cryptography[M]. Boca Raton, Florida: CRC Press, 1997.
- [8] Lancaster P, Tismenetsky M. The Theory of Matrix (2nd ed.)[M]. Orlando, Florida: Academic Press, 1985.
- [9] Castleman K R. Digital Image Processing (2nd ed.)[M]. Englewood Cliffs, New Jersey: Prentice Hall, 1997.

作者简介:



赵险峰 男, 1969年6月出生于安徽淮北市, 分别于1991年和1999年于大连理工大学和南昌大学获得学士学位和硕士学位, 现为上海交通大学计算机科学与工程系博士研究生, 主要研究兴趣包括多媒体处理和信息安全。



汪为农 男, 1949年8月出生于安徽巢湖市, 1991年于柏林工业大学和上海交通大学计算机科学与工程系获得博士学位(联合培养), 现为上海交通大学计算机科学与工程系教授、博士生导师、中国教育科研网专家组成员, 主要研究方向为网络安全和信息安全。