基于感兴趣区域的图像认证技术

张小华 孟红云* 刘 芳** 焦李成
(西安电子科技大学智能信息处理研究所 西安 710071)
*(西安电子科技大学理学院 西安 710071)
**(西安电子科技大学计算机学院 西安 710071)

摘 要：信息认证是网络安全的一个重要方面，其目的在于判断信息的完整性和可信性。该文在对以往的图像认证技术分析的基础之上，给出了一种基于感兴趣区域的图像认证算法。首先将要认证的图像分割成一定大小的基本图像块，然后根据用户给出的感兴趣区域，后续分割每一个基本图像块，并提取相应各子块的签名信息并保存分割附加码，最后将签名信息隐藏于图像子块的中频系数中。实验结果证明算法不仅对偶然攻击具有较强的鲁棒性，而且对恶意攻击具有较高的检测、定位能力,同时具有较强的自适应能力。

关键字：图像认证，数字水印技术，感兴趣区域，图像分割
中图分类号：TP391 **文献标识码**：A **文章编号**：1009-5896(2004)01-0031-04

Effective Image Authentication Technique Based on the Interested Region

Zhang Xiao-hua Meng Hong-yun* Liu Fang** Jiao Li-cheng
(*Xidian University, Xi'an* 710071, *China*)
*(*Dept of Math.*, *Xidian University, Xi'an* 710071, *China*)
**(*Dept of Computer*, *Xidian University, Xi'an* 710071, *China*)

Abstract Information authentication technique is an important inspects of network security, and its motive lies in authentication of the integrity and creditability of information. Based on the analysis of the previous image authentication techniques, a new effective image authentication technique is proposed based on the interested region. Firstly, the original image is divided into a certain size preliminary image blocks. Then on the basis of the interested region given by users, the preliminary image block is divided into smaller blocks, at the same time the signature of each image block is extracted.Finally the signature of image blocks is embedded into intermediate frequency coefficient of image blocks. Experimental results are provided to demonstrate the effectiveness, self-adaptability, and validity of the method.

Key words Image authentication, Digital watermarking technique, The interested region, Image segment

1 引言

多媒体信息认证技术是目前信息认证技术研究的热点，相对经典需要认证的秘密信息，多媒体信息具有数据量大、冗余性高、保密性低等特性。数字图像认证技术是多媒体信息认证研究的重要方向之一。

目前大多数图像认证技术并没有考虑所要保护图像的内容，且认为构成图像的所有区域在传输过程中所遭遇攻击是相同的。JPEG2000[1]成为第一个基于感兴趣区域的压缩技术，在压缩过程中针对不同区域采用不同的方法和压缩强度。事实上，在实际应用中观察者对图像的各个部分的关注程度并不相同，关注程度高的称为感兴趣区域，反之，称为非感兴趣区域。本文认为感兴趣区域受偶然攻击的概率低于非感兴趣区域，而受恶意攻击的概率大于非感兴趣区域。基于此考虑，本文给出了一种基于感兴趣区域的图像认证技术。

2 基于感兴趣区域的图像分割算法

为了使得认证算法具有一定的定位能力，往往需要将宿主图像分割为独立的子块。显然图像子块尺寸越小，定位的精度越高，但计算量却越大，所以需在定位精度和计算量之间进行折衷。记原始图像为 $I(i,j)$，感兴趣区域为 $\mathrm{ROI}=\{\mathrm{ROI}^r(m,n),r=1,\cdots,R\}$，$R$ 为感兴趣区域的总数目。为了在认证端，能精确重现分割过程，我们采用分割附加码 AddCode 来记录分割过程。具体的分割算法如下：

2003-03-24 收到，2003-10-28 改回
国家自然科学基金(60073053, 60133010)资助课题

(1)将原始图像分割为大小为 $P\times Q$ 的基本图像块 SubI^l，取 $k=1$，并构造待分感兴趣序列 $\mathrm{Sub}_{\mathrm{roi}}$，初始长度 $\mathrm{len}=0$，然后判断 $\mathrm{SubI}^l \cap \mathrm{ROI}$ 是否为空集，如果等于空集，$\mathrm{AddCode}(k)=1$，$k=k+1$，继续判断 $\mathrm{SubI}^{l+1} \cap \mathrm{ROI}$；如果不等于空集，$\mathrm{AddCode}(k)=0$，$k=k+1$，将 SubI^l 插入 $\mathrm{Sub}_{\mathrm{roi}}$，$\mathrm{len}=\mathrm{len}+1$。

(2)从 $\mathrm{Sub}_{\mathrm{roi}}$ 中抽取一感兴趣图像块 SubI^l，$\mathrm{len}=\mathrm{len}-1$，并将它一分为四：$\mathrm{SubI}_h^l$，$h=1,\cdots,4$，然后判断每一个 $\mathrm{SubI}_h^l \cap \mathrm{ROI}$ 是否为空集，如果等于空集，$\mathrm{AddCode}(k)=1$，$k=k+1$；如果不等于空集，$\mathrm{AddCode}(k)=0$，$k=k+1$，然后判断 SubI_h^l 的大小是否等于最小块 $U\times V$，不等于，则将 SubI_h^l 插入 $\mathrm{Sub}_{\mathrm{roi}}$，$\mathrm{len}=\mathrm{len}+1$。重复（2），直到 $\mathrm{len}=0$。

计算量是衡量一个算法有效性的重要标志，本文以判断 $\mathrm{SubI}_h^l \cap \mathrm{ROI}$ 是否为空集的次数 Num 作为分割算法的计算量的标志。为了计算方便我们假设基本图像块 SubI^l 的大小为 $2^S * 2^S$，最小图像块的大小 $U=V=2^r$，则 Num 的最大值和最小值分别为

$\mathrm{Num}_{\max}=L\times(1+4+4^2+\cdots 4^{S-r})\times R=L\times(4^{S-r+1}-1)\times R/3$。

$\mathrm{Num}_{\min}=(L+4\times(S-r))\times R$。其中 L 为基本图像块的数目，R 为感兴趣区域的数目。

3 水印签名信息提取与嵌入

基于宿主图像内容的签名水印信息的优点在于水印信息来自于图像本身，攻击者很难通过直接复制可信水印信息进行攻击[2]。Lin[2]以两个图像块的DCT变换的同一位置的两个系数之间的大小关系作为签名水印信息，Bhattacharjee等人[3]以特征因子点位置的数字签名作为水印信息，Queluz[4]以图像块的矩和边缘作为签名水印信息。用于图像认证的签名水印信息，应满足：具有较强的鲁棒性，可区分偶然攻击和恶意攻击；具有代表性，签名水印信息应是图像重要特征的反映，判断图像是否变化等价于签名是否变化。

3.1 签名水印提取算法

假设图像块的特征比特 $f\in\{0,1\}^n$，签名信息 $T\in\{0,1\}^m$，则签名提取函数可定义为 $F:\{0,1\}^n\to\{0,1\}^m$。本文以图像块 I_{sub} 的DCT或DWT变换的低频系数作为该图像块的特征系数，记为 $s_{\mathrm{low}}=(s_1,\cdots,s_L)$，由于 $(s_1,\cdots,s_L)\in R^L$，所以在提取签名之前须将 s_{low} 转换为整数。本文给出了如下量化转换方法：

$(y_1,\cdots,y_L)=(Q_\Delta(s_1),\cdots,Q_\Delta(s_L))$，其中 $Q_\Delta(s_l)=\mathrm{round}\left(\frac{s_l}{\Delta}\right)$，

Δ 为量化步长。将 y_l 用二进制表示并连接起来，即可得到图像块 I_{sub} 的特征比特 $f=(f_1,\ldots,f_n)\in\{0,1\}^n$。作为一个有效的签名提取函数必须满足：当 f 发生变化时，$F(f)$ 也应发生变化，$F(f)$ 变化的概率和 f 变化的强度成正比，同时应满足不可逆性。本文给出如下的签名提取函数：

$F(f)=(F_1(f),\cdots,F_m(f))$，其中 $F_i(f)=\mathrm{major}(f_1,\cdots,f_{2N+1})$，其中 $(f_1,\cdots,f_{2N+1})$ 是从 $(f_1,\cdots,f_n)$ 中基于密钥 K 随机选出的 $2N+1$ 个特征比特。若记 $(f_1,\cdots,f_{2N+1})$ 中出现0的个数为 n_0，出现1的个数为 n_1，则主要比特提取函数 $\mathrm{major}(f_1,\cdots,f_{2N+1})$ 定义为

$$F_i=\mathrm{major}(f_1,\cdots,f_{2N+1})=\begin{cases}1, & n_1>n_0\\ 0, & \text{其他}\end{cases} \tag{1}$$

为了使所构造的 F_i 具有一定的鲁棒性，本文以 $F(s_l)=s_l+\alpha\times(Q_\Delta(s_l)\times\Delta-s_l)$ 代替 s_l，其中 α 为稳定参数，当 I_{sub} 属于感兴趣区域时，$\alpha=1-2w$；当 I_{sub} 属于非感兴趣区域时，$\alpha=1-w$。这是因为在传输过程中，偶然攻击一般会尽可能避免破坏感兴趣区域的质量，而同时为了提高压缩比，不得不去破坏非感兴趣区域的质量。同时本文在提取签名信息时根据图像块的类型提取不同长度的签名信息，使得从感兴趣区域提取的签名长度 L_{roi} 大于从非感兴趣区域所提取的签名的长度 L_{not}。

3.2 签名构造算法性能分析

主要比特提取函数 major 对于 $(f_1,\cdots,f_{2N})$ 变化的敏感性对于本文图像认证算法的性能将起到至关重要的作用。下面我们将讨论当 $(f_1,\cdots,f_{2N})$ 中有 d 比特发生变化，主要比特随之变化的概率 $P(d)$。假设 $\mathrm{sum}=f_1+\cdots+f_{2N+1}$，显然 sum 等于 k 概率为 $C_{2N+1}^k/2^{2N+1}$，则概率 $P(d)$ 的推导过程如下：

假设 $(f_1,\cdots,f_{2N+1})$ 中的 $(f_{k_1},\cdots,f_{k_d})$ 发生变化，发生变化的0的个数为 x，1的个数为 y。假设 $f_1+f_2+\cdots+f_{2N+1}=k$，分为两种情况讨论：签名由0变为1，由1变为0。

很显然，要使所提取的签名比特发生变化，则 x 和 y 必须满足条件

$$\left.\begin{array}{l}x+y=d\\ k-y+x\ge N+1\\ x\le 2N+1-k\\ y\le k\\ d\ge x,y\ge 0\\ k\le N\end{array}\right\}\ \text{或}\ \left.\begin{array}{l}x+y=d\\ k-y+x\le N\\ x\le 2N+1-k\\ y\le k\\ d\ge x,y\ge 0\\ k\ge N+1\end{array}\right\} \tag{2}$$

对于第一种情况即签名由0变为1，则 y 可取的最大和最小值为

$Y_{\mathrm{up}}=\min(k,\lfloor(k+d-N-1)/2\rfloor,d)$，$Y_{\mathrm{low}}=\max(d+k-2N-1,0)$，$\lfloor\cdot\rfloor$ 为向下取整函数。而 k 可取的最小值为 $K_{\mathrm{low}}=\max(0,N+1-d)$，最大值为 N，则 $(f_{k_1},\cdots,f_{k_d})$ 变化前后所构造的主要比特发生变化的概率为

$$P_1(d)=\Big(\sum_{k=K_{\mathrm{low}}}^{N}p(\mathrm{sum}=k)\sum_{y=Y_{\mathrm{low}}}^{Y_{\mathrm{up}}}C_k^y * C_{2N+1-k}^{d-y}\Big)/C_{2N+1}^d \tag{3}$$

对于第二种情况即签名由1变为0，则 y 可取的最大和最小值为 $Y_{\mathrm{up}}=\min(k,d)$，$Y_{\mathrm{low}}=\max\lfloor(k+d-N)/2\rfloor,0,d+k-2N$

$-1)$，$\lfloor . \rfloor$ 为下取整函数，而 k 的最大值为 $K_{up}=\min(2N+1, N+d)$，最小值为 $N+1$，则 $(f_{k_1},\cdots,f_{k_d})$ 变化前后所构造的主要比特发生变化的概率为

$$P_u(d)=\left(\sum_{k=N+1}^{K_{up}} p(\text{sum}=k)\sum_{y=Y_{low}}^{Y_{up}} C_k^y * C_{2N+1-k}^{d-y}\right)\Big/ C_{2N+1}^{d} \quad (4)$$

因此，当 $(f_1,\cdots,f_{2N+1})$ 中有 d 比特发生变化，所提取的签名比特发生变化的概率 $P(d)$：

$$P(d)=p_l(d)+p_u(d) \quad (5)$$

以 N=64为例，图1给出 d 与 $p(d)$ 的关系图。容易看出，当 d 越大，$P(d)$ 越大；当所有的 f_k 都发生变化，则提取的签名肯定变化。

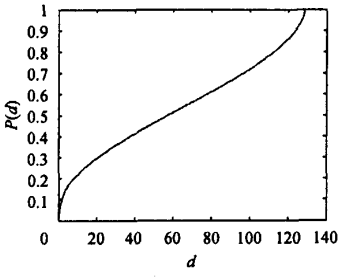

图1　$P(d)$与d关系图

3.3 签名信息嵌入算法

对于宿主信号 x_n 满足独立同分布（IID），干扰为加性高斯白噪声（AWGN）的数字水印隐藏问题，Costa[5]给出理论上最优数字水印算法，但由于该算法涉及到构造一个非常庞大的码本，所以在实际中并没有得到广泛的使用，直到Chen[6]给出一个次最优码本：

$$U=\{u=(l+k_n)\alpha\Delta+(d_N/2)\alpha\Delta \,|\, k_n\in(0,1), d_n\in\{0,1\}, l\in Z\}$$

所谓的将水印信息 $d_n\in\{0,1\}$ 嵌入到宿主信号 x_n 中，就是从集合 U 中选取一元素 u_x 使得 $|x_n-u_x/\alpha|$ 最小，很显然，$u_x=\alpha Q_\Delta\left(x_n-\Delta\left(d_n/2+k_n\right)\right)+\alpha\Delta\left(d_n/2+k_n\right)$，其中 $Q_\Delta(x)$ 为 x 关于步长 Δ 的一致量化。则信号：

$$s_n=x_n-\alpha\left(Q_\Delta\left(x_n-\Delta\left(d_n/2+k_n\right)\right)-x_n+\Delta\left(d_n/2+k_n\right)\right)$$

隐含水印信息 d_n。

由上可知，经过基于感兴趣区域的图像分割后，原始图像 $I(i,j)$ 被分为非感兴趣块序列 $I^r_{\text{Not}}, r=1,\ldots,R_{\text{not}}$ 和感兴趣块序列 $I^r_{\text{roi}}, r=1,\ldots,R_{\text{roi}}$，为了保护感兴趣区域质量和提高其抗干扰能力，我们采用如下嵌入策略：

(1)将从 I^r_{roi} 中提取的签名 $T^r_{\text{roi}}(i), i=1,\ldots,L_{\text{roi}}$，经混沌加密后，通过Costa方法，嵌入到对应 I^r_{roi}，其中 $\alpha_{\text{roi}}=1-2*w$。

(2)将从 I^r_{not} 中提取的签名 $T^r_{\text{not}}(i), i=1,\ldots,L_{\text{not}}$，经混沌加密后，通过Costa方法，嵌入到对应 I^r_{not}，其中 $\alpha_{\text{not}}=1-w$。当图像接收者，不知道感兴趣区域的位置时，就必须附加传输分割附加码，当然这会为认证带来一定的额外工作量。解决的方法有二：方法一，双方约定，那些内容是感兴趣内容，且将包含感兴趣内容的最小区域作为感兴趣区域；方法二，将隐含签名水印的图像分割为大小相同的图像块，然后将分割附加码或感兴趣区域的位置信息隐藏其中。一般情况下，这些信息的信息量较小，所以他们的嵌入并不会破坏其中所隐含的签名水印信息。

4　签名信息认证和性能分析

图像在传输过程中会遭遇偶然攻击或恶意攻击，所以当接收者接收到图像后，需要认证其可信性和完整性。本文认证算法的认证过程分为当前图像签名提取、隐含签名水印提取、签名比较。在构造签名前，首先要根据双方约定或者当前图像所隐含的分割附加码 AddCode，重新分割图像，然后根据给出的签名水印构造算法，即可获得当前图像的签名信息 T''_{not} 和 T''_{roi}。假设用户接收到隐含签名的系数为 $s'_n=s_n+v_n$，其中 v_n 为加性噪声干扰，若记 $y_n=Q_\Delta(s'_n-k_n\Delta)-(s'_n-k_n\Delta)$，我们给出如下的解码判断规则：

$$d'_n=\begin{cases}0, & |y_n|-\Delta/4<0\\ 1, & \text{其他}\end{cases} \quad (6)$$

这样就可得到其中所隐含签名信息 T'_{not} 和 T'_{roi}，最后比较当前签名信息和其中所隐含的签名信息，如果两者相匹配，表明图像块内容没有变化，是可信的；否则，不可信；

为了证明算法的有效性，我们分别从图像的失真、抗干扰能力、定位精度以及安全性等 4 方面分析该算法。本文主要分析签名隐藏所造成图像失真，记误差为 e_n：

$e_n=x_n-s_n=-\alpha\times(Q_\Delta(x_n-m_n)-(x_n-m_n))$，则 e_n 的期望 $E(e_n)$ 和方差 σ^2 为 $E(e_n)=0$，$\sigma^2=\alpha\times\Delta^2/12$，显然感兴趣区域失真较小。定义集和 U_{d,k_n}：$U_{d,k_n}=\{s_n \,|\, s_n=x_n+\alpha\times(Q_\Delta(x_n-(k_n+d/2)\times\Delta)-x_n-(k_n+d/2)\times\Delta)\}$，则抗干扰直径可定义为 $D=\min|u_1-u_2|$，其中 $u_1\in U_{d_1,k_n}$，$u_2\in U_{d_2,k_n}$，且 $d_1\neq d_2$。抗干扰直径 D 的作用在于描述不同码集之间的距离，衡量算法抵抗干扰的能力，D 越大抵抗偶然攻击的能力就越大，对修改的敏感性就越弱，本文算法的抗干扰直径 $D=(2\alpha-1)\Delta/2$。对于非感兴趣区域本文的最好定位精度可达 $U\times V$ 图像块，而最差为 $P*Q$，而对于感兴趣区域定位精度可达 $U\times V$ 的图像块，本文算法的安全性主要体现在如下两方面：

(1)在隐藏签名信息之前，通过混沌序列 C_n 产生加密签名 T_n，加密方法为 $d_n=C_n\oplus T_n$，解密方法为 $T_n=d_n\oplus C_n$，"$\oplus$"为"或"运算。

(2)通过混沌迭代生成状态信息 k_n，加密 s_n。

5　实验结果与分析

为了证明算法的有效性，我们进行了如下实验。以大小为 256*256 灰度图像作为测试对象，基本图像块大小为 64×64，分割最小块 $U\times V=8\times8$，感兴趣区域数目为 3，且

以 logistic 混沌序列的初始参数作为密钥 K。图 2 为原始图像，白色矩形包围区域为感兴趣区域，图 3 为基于感兴趣区域分割算法的分割结果。图 4 为隐含签名信息和分割附加码的宿主图像，相应的 PSNR=44.1352。图 5 受高斯噪声干扰后的图像，相应的均值和方差为 $\mu=0,\sigma=5$，PSNR=26.6273。图 6 为压缩因子 $\eta=20\%$的压缩图像。图 7 为被篡改的图像。图 8、图 9、图 1 0 为图 5、图 6、图 7 的认证结果，白色图像块为签名不匹配的区域。当噪声方差小于或等于 5 时，极少出现不匹配区域，而当方差大于 5 时会出现一部分不匹配区域，这些区域中一部分集中在感兴趣区域。原因有三：其一，噪声方差较大，噪声强度出现较大值的概率增加,导致小区域低频系数变化较大；其二，非感兴趣图像块的尺寸一般较大，所以噪声对大区域块的低、中频信息平均影响较小；其三，非感兴趣区域抗干扰直径较大。当 JPEG 压缩因子小于 20%时，会在感兴趣区域出现一些签名不匹配的图像块，而对于 JPEG2000 可以压缩到更高的压缩比，却不会出现不匹配区域，因为在 JPEG2000，为了提高整体压缩比，会尽可能压缩非感兴趣区域，保护感兴趣区域，而本文算法的非感兴趣区域的抗干扰直径大于感兴趣区域。由于感兴趣区域对篡改的敏感度强于非感兴趣区域，所以用户会很容易发现被篡改的感兴趣区域，同时由于感兴趣区域的图像块较小，所以检测定位的精度相对较高,如图 10。以上实验结果证明该算法是有效可行的。

图 2 原始图像及感兴趣区域

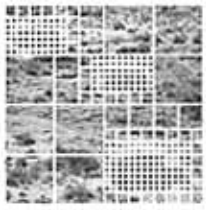

图 3 分割结果

图 4 含有签名的可认证图像

图 5 加噪图像

图 6 JPEG 压缩图像

图 7 被篡改图像

图 8 加噪认证结果

图 9 JPEG 压缩图像

图 10 篡改图像认证结果的认证结果

参 考 文 献

[1] ISO/IEC JTC 1/SC 29/WG 1 ISO/IEC FCD 15444-1: Coding of Still Pictures. Mar, 2000, http://www.jpeg.org/FCD15444-1.htm.

[2] Lin Ching-Yung, Chang Shih-Fu. A robust image authentication method distinguishing JPEG compression from malicious manipulation. *IEEE Trans. on Circuits and Systems of Video Technology*, 2001, 1(2): 153 – 168.

[3] Bhattacharjee S, Kutter M. Compression tolerant image authentication, Proc. IEEE Int. Conf. Image Processing, Chicago, 1998, vol.1: 435 – 439.

[4] Queluz M P. Content-based integrity protection of digital images, SPIE Conf. Security and Watermarking of Multimedia, San Jose, 1999, Vol.3657: 85 – 93.

[5] Costa M H M. Writing on dirty paper. *IEEE Trans. on Information Theory*, 1983, 29(3): 439 – 441.

[6] Chen B,Wornell G. Preprocessed and post processed quantization index modulation methods for digital watermarking, Proc. of SPIE, San Jose, 2000, Vol.3971: 48 – 59.

张小华： 1974 年生，博士生，研究方向为图像处理、信息安全、进化算法.

孟红云： 1975 年生，博士生，主要从事进化算法、数字水印方面研究.

刘 芳： 1963 年生，副教授，研究方向智能信息处理、模式识别、电子商务.

焦李成： 1959 年生，教授，博士生导师，研究方向，智能信息处理、非线性理论、数字水印.