一种基于曝光度的数字水印攻击方法

李开拓[1)]　　张亶[2)]

[1)](浙江大学软件学院,杭州　310027)　[2)](浙江大学计算机科学与技术学院,杭州　310027)

摘　要：本文提出了一种针对 JPEG 压缩域嵌入算法 SWE 和 IWE 的水印攻击方法。首先本文回顾了 SWE 案和 IWE 算法，然后分析了曝光度和图像亮度的关系以及曝光度的改变对水印的检测和提取所带来的干扰。在此基础上，提出了一种基于曝光度的水印攻击方法。该方法通过对水印图像进行不同的曝光度处理，然后将处理后的若干幅图像做加权平均，得到合成图像。实验证明，该攻击方法简单有效，从合成后的图像中无法正常提取水印，并且合成图像保持了良好的主观质量。

关键词：曝光度；JPEG 压缩；水印攻击

Attack on Digital Watermarking Based on Exposure

LI Kai-tuo[1)], ZHANG Dan[2)]

1) (*Software College, Zhejiang University, Hangzhou 310027*)2) 2)(*College of Computer Science and Technology, Zhejiang University, Hangzhou 310027*)

Abstract: In this paper, a simple and effective attack on the blind watermarking techniques in JPEG compressed domain—SWE and IWE are proposed. After reviewing the algorithms of the SWE and IWE, it is discovered that the SWE and IWE watermarking scheme can not derive correct decoded watermark when the watermarked image is exposed differently. Therefore, a watermark attack based on multi-exposed images is proposed, which can pass through the watermark extraction and watermark authentication successfully. This attack scheme demonstrates that the watermarking algorithms are flawed which can be used by the malicious content owner to forge and distribute a watermarked image to the unknown users. At last, an experiment of watermark attack based on multi-exposed images on SWE and IWE watermarking is devised and the results show that our attack is successful with a reasonable image quality.

Keywords: exposure; JPEG compression; watermark attack

1 引 言

数字水印能够用于对多媒体数据进行版权保护和内容认证。自从上个世纪九十年代中期以来，数字水印一直是人们研究的热点领域，各种水印算法层出不穷[1]。许多文献提出压缩域数字水印嵌入技术[2~5]，即直接将水印嵌入到压缩位流或索引中。基于 JPEG 标准的压缩域的方法是其中一种。这类方法不仅能够有效抵抗 JPEG 压缩攻击，而且随着 JPEG 文件格式的广泛应用而具有很大的实用价值。文献[5]提出了两种 JPEG 压缩域嵌入水印的模型：第一种是 JPEG-to-JPEG 水印模型；另外一种是 JPEG2000-to-JPEG2000 水印模型。依据第一种模型，文献[5]给出了一种具体的量化嵌入方案 SWE 和在 SWE 基础上做迭代嵌入的方案 IWE。本文利用曝光度和图像亮度之间的关系，提出一种针对 SWE 和 IWE 的攻击方法。该方法通过对水印图像进行不同的曝光度处理，然后将处理后的若干幅图像做加权平均，最后得到的图像保持了良好的主观质量。通过实验可以表明，该攻击方法有效的破坏了水印的提取。

基金项目：国家“973”重点基础研究发展规划资助项目（2006CB303104）。

第一作者简介：李开拓（1985~ ），男，浙江大学软件学院硕士研究生。研究方向为数字图像处理。
E-mail: lee.kaituo@gmail.com。

2 SWE 和 IWE 简介

2.1 SWE 简介

待嵌入水印的序列来自于 JPEG 文件霍夫曼解码和解量化后，按照 zig-zag 的顺序，以一定的比例，从各个 8×8 的 DCT 块中相应位置取出的交流系数的集合，这个集合被称作水印宿主矢量， 用 $Y=[y_1,y_2,\cdots,y_M]$ 表示。水印序列 $W=[w_1,w_2,\cdots,w_N]$ 为二值序列，其中 $M \gg N$ 。为了在 M 个系数中嵌入 N 比特，先把向量 Y 分成 N 个子矢量 Y_i，每个子矢量的长度为 $P=\lfloor M/N \rfloor$。每个子矢量嵌入一比特的水印信息。SWE 采用了两个私钥，第一个私钥 $D=[d_1,d_2,\cdots,d_N \mid d_i \in R^+, 1\le i\le N]$ 是 N 个伪随机的正实数集，第二个私钥 $K=[k_1,k_2,\cdots,k_N]$ 中每个 k_i 都服从均值为 0 的高斯分布。SWE 用这两个私钥将水印 W 嵌入 Y，然后根据 K 和 D 提取水印。嵌入水印后的 Y 用 Y' 表示。基于以上定义，SWE 采用的基本嵌入操作如下：

$$Y_i' = Y_i + \alpha_i K_i \tag{1}$$

其中

$$\alpha_i=\begin{cases}\dfrac{d_i\cdot round(\langle X_i,K_i\rangle/d_i)-\langle X_i,K_i\rangle}{\|K_i\|_2^2} & \text{情况1}\\ \dfrac{d_i\cdot[round(\langle X_i,K_i\rangle/d_i)+1]-\langle X_i,K_i\rangle}{\|K_i\|_2^2} & \text{情况2}\\ \dfrac{d_i\cdot[round(\langle X_i,K_i\rangle/d_i)-1]-\langle X_i,K_i\rangle}{\|K_i\|_2^2} & \text{情况3}\end{cases} \tag{2}$$

情况 1：当 $[round(\langle X_i,K_i\rangle/d_i)]\bmod 2 = w_i$

情况 2：当 $[round(\langle X_i,K_i\rangle/d_i)]\bmod 2 \ne w_i$ ，且 $\langle X_i,K_i\rangle \ge d_i\cdot round(\langle X_i,K_i\rangle/d_i)$

情况 3：当 $[round(\langle X_i,K_i\rangle/d_i)]\bmod 2 \ne w_i$ ，且 $\langle X_i,K_i\rangle < d_i\cdot round(\langle X_i,K_i\rangle/d_i)$

在提取水印时，替该嵌入操作对应的盲提取操作如下：

$$w_i' = round(\langle Y_i',K_i\rangle/d_i)\%2 \tag{3}$$

最后根据相关系数 s1 的大小评价恢复出的水印信息和嵌入的信息的相关程度：

$$s1=\frac{\sum_{i=1}^{N}(2\cdot w_i-1)\cdot(2\cdot w_i'-1)}{\sqrt{\sum_{i=1}^{N}(2\cdot w_i-1)^2\cdot\sum_{i=1}^{N}(2\cdot w_i'-1)^2}} \tag{4}$$

2.2 IWE 简介

IWE 是为了弥补 SWE 对 JPEG 重量化攻击缺乏鲁棒性的弱点而提出的。在基本嵌入操作之前，将 Y_i 与一个长度为 P 的随机噪声子矢量 N_i 相加。N_i 的每个元素服从 $[-q/2, q/2]$ 上的均匀分布，其中 q 是 Y_{ij} 所对应的量化阶距，它同时考虑了量化表和嵌入因子。然后对修改后的 Y_i 进行基本嵌入操作，并用原始图像的量化表和质量因子进行量化，继而做编码压缩。接着再解码，提取出水印，如果相关系数大于阈值，在算法停止，否则重复上述过程。

2.3 评价

文献[5]用各种流行的攻击方法检验 SWE 和 IWE 的鲁棒性，结果如下：SWE 和 IWE 对 JPEG 压缩攻击、中值滤波攻击、嵌入水印攻击、尺度变换攻击、行列去除攻击、旋转加裁剪攻击、旋转加尺度变换攻击、自相似攻击、平移攻击、局部弯曲攻击具有鲁棒性；对仿射变换攻击、高斯滤波攻击、锐化攻击、裁剪攻击、噪声叠加攻击、大幅度旋转攻击不具备鲁棒性。但是第二种情况里面的攻击是以导致图像质量的严重下降为代价的。因此，SWE 和 IWE 能够确保对绝大多数不至于破坏水印作品的攻击具有鲁棒性。IWE 算法的流程如图 1 所示。

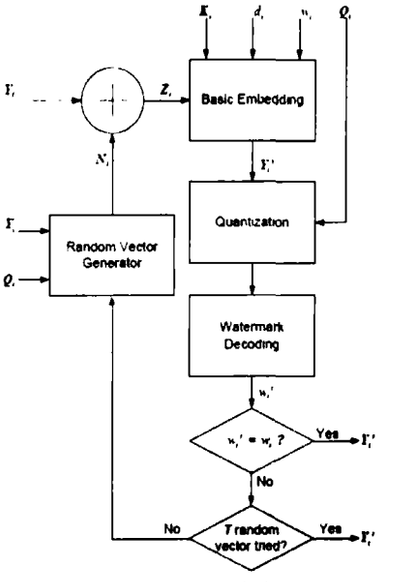

图 1 IWE 算法流程

Fig.1 The iteration loop in IWE

2.4 实验结果

图 2~图 5 是 SWE 及 IWE 算法的实验结果。实验中的原始图像“lena”为 512×512 的灰度图像，如图 2 所示。水印图像为 40×40 的二值水印图像，如图 3 所示。水印的嵌入和提取的实验结果

如图 4、5 所示，可以看出原始图像和嵌入水印后的图像基本上没有区别，表明了 SWE 及 IWE 算法具有良好的主观图像质量。

图 2 原始图像
Fig.2 Original "Lena"

图 3 水印图像
Fig 3. Watermark

图 4 嵌入水印的图像
Fig.4 IWE-watermarked image

图 5 提取出来的水印
Fig.5 Decoded watermark

3 曝光度和水印

3.1 曝光度的背景

曝光度是摄影中的概念，指的是在拍照时落在摄影媒介（比如 CCD）上的光粒子数量。照片的好坏与曝光度有关，也就是说应该用多少的光粒子使 CCD 能够得到主观图像质量好的图像。曝光度与通光时间（快门速度决定）、通光面积（光圈大小决定）有关。

之所以选择曝光度作为水印攻击的切入点，不仅因为曝光度的改变有效的干扰了水印图像的提取，而且因为不同曝光度的图像是在自然状态下真实可得的图像，这样得到的图像具有良好的主观质量。这样我们可以利用曝光度的改变作为攻击手段，破坏水印的功用，同时还能保证图像质量。

3.2 曝光度和图像亮度的关系

一般说来，一幅数字图像从 CCD 的颜色空间变到 YCbCr 颜色空间下需要经过三个步骤[6]：

（1）通过与矩阵相乘，从 CCD 的颜色空间变到 sRGB（线性）空间。

（2）通过使用 Gamma 补偿，从 sRGB（线性）空间变到 sRGB（非线性）空间。

（3）通过与矩阵相乘，从 sRGB（非线性）空间变到 YCbCr 空间。

因为与矩阵相乘是线性变换，所以它不会破坏像素值和曝光度之间的线性关系。但是 Gamma 补偿是非线性变换，所以会破坏上述线性关系。在这里 gamma 函数可以表示为[7]：

$$V_{out} = V_{in}^{\gamma} \qquad (5)$$

通常 γ 取 0.45，这样可以得到曝光度改变量和 Y 分量改变量的关系：

$$Y = Y_0(\frac{EV}{EV_0})^{0.45} \qquad (6)$$

其中，改变前的图像的曝光度为 EV_0，Y 分量的值为 Y_0，改变后的图像的曝光度为 EV，Y 分量的值为 Y。比如说，如果曝光度增大两倍，那么 Y 分量的值会增大 $Y/Y_0 = (2)^{0.45} = 1.366$ 倍。

3.3 改变曝光度对水印提取和检测的影响

依据式（6），改变曝光度将对 Y 分量产生一个非线性的扰动。同时，下面的实验结果将证明改变曝光度对水印提取和检测都将有较大的影响。为简单起见，只考虑灰度图像。曝光度的放大和缩小对 s1 以及检测错误率的影响如图 6 所示。曝光度的缩放和 PSNR 的关系如图 7 所示。可以看出曝光度的缩放程度越大，对水印的检测和提取所带来的干扰越大。

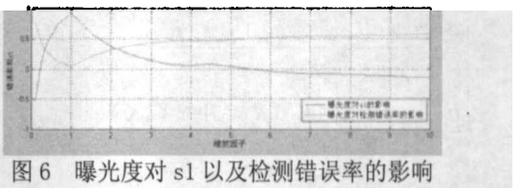

图 6 曝光度对 s1 以及检测错误率的影响
Fig.6 The influence of exposure on s1 and error rate

图 7 曝光度对 PSNR 的影响
Fig7 The influence of exposure on PSNR

4 基于曝光度的水印攻击

4.1 加权平均函数的引入

如上文所述，对于经过曝光处理的嵌入水印图像，随着曝光度的改变量越来越大，使用水印检测器检测水印就越来越难，但是对嵌入水印的图像的质量的损害也越来越明显。由此本文将多幅经过不同曝光处理的图像做加权平均，使得经过改动的图像仍具有很高的逼真度，同时检测器无法检测出水

印的存在。对于曝光度改变量越大的嵌入水印图像，其权值越高。假设有 k 幅不同曝光度的图像 $X=\{x_1,x_2,\cdots,x_k\}$，它们对应的权值为 $\lambda=\{\lambda_1,\lambda_2,\cdots,\lambda_k\}$，则攻击中采用的函数为：

$$x^{wave}=\sum_{j=1}^{k}\lambda_j x_j \Big/ k \tag{7}$$

4.2 实验结果

根据（7）式，对实验图像进行了基于曝光度的水印攻击。不同曝光度的图像件数对水印检测错误率和 s1 的影响如图 8 所示。不同曝光度的图像件数和 PSNR 的关系如图 9 所示。可以看出随着图像件数的增多，水印检测的错误率也将上升，同时引入的失真在可接受的范围之内。图 10(a~d)分别为图像件数为 11、41、81、121 的合成图像，图 10（e~h）为从对应的图像中提取出的水印。

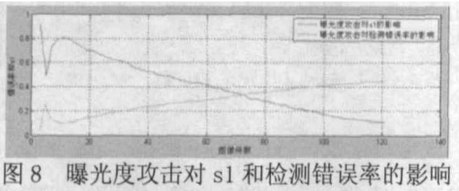

图 8 曝光度攻击对 s1 和检测错误率的影响

Fig.8 The influence of attack based on exposure on s1 and error rate

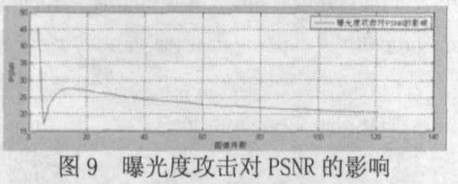

图 9 曝光度攻击对 PSNR 的影响

Fig.9 The influence of attack based on exposure on PSNR

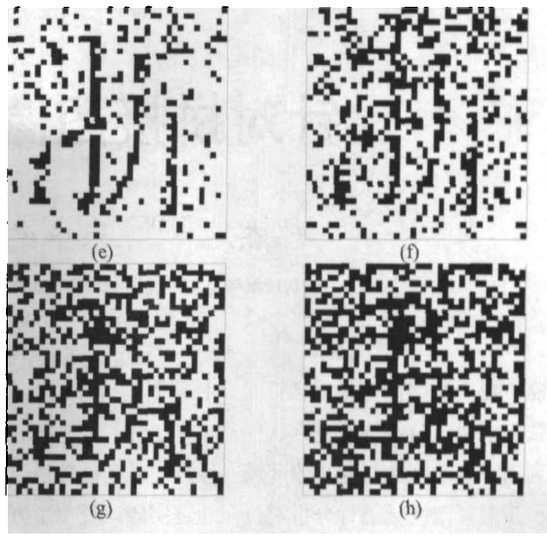

图 10 合成图像和提取出来的水印

Fig.10 Composite image and extracted watermark

5 结束语

本文提出一种利用曝光度的数字水印攻击方法。主要思想为：将原始的嵌入水印的图像进行不同的曝光处理，然后对这些图像做加权平均，最后得到合成图像。实验验证了方法的有效性，合成图像仍具有很高的逼真度，同时检测器无法检测出水印的存在。下一步的工作是设计一种更加有效的图像合成方法，在确保移除水印的同时，进一步提高合成图像的质量。

参考文献

[1]I. J. Cox, M. L. Miller, J. A. Bloom. Digital Watermarking[M], San Francisco, San Diego, CA, USA: Morgan Kaufmann Publishers, 2001.

[2]M. Iwata, K. Miyake, A. Shiozaki. Digital Watermarking Method to Embed Index Data into JPEG Images[J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, October, 2002, E85-A(10): 2267~2271.

[3]Suhail MA, Obaidat M S. Digital watermarking-based DCT and JPEG model[J]. IEEE Trans. on Instrumentation and Measurement, 2003, 52(5): 1640~1647.

[4]J. Fridrich, M. Goljan, R. Du. Invertible authentication watermark for JPEG images[A]. In: Proc. ITCC[C], Las Vegas, Nev, USA, April 2001.

[5]P.H.W. Wong. Image Watermarking and Data Hiding Techniques[D]. Hong kong: The Hong Kong University of Science and Technology, 2003.

[6]D.S. Taubman, M.W. Marcellin. JPEG2000: Image Compression Fundamentals, Standards, and Practice[M], Norwell, MA, USA: Kluwer Academic Publishers, 2001.

[7]Charles A. Poynton. Digital Video and HDTV: Algorithms and Interfaces[M], San Francisco, San Diego, CA, USA: Morgan Kaufmann Publishers, 2003.