基于 ICA 的抗拷贝攻击的数字水印方案

胡慧博 刘 琚 孙建德
(山东大学信息科学与工程学院 济南 250100)

摘 要：该文提出了一种基于独立分量分析(Independent Component Analysis, ICA)的能有效抵抗拷贝攻击的数字水印方案。该方案将能够唯一描述原始图像的图像签名与数字水印一同嵌入到原始图像中，并采用 ICA 技术进行检测。通过验证签名，可以判断待测图像是否合法，从而实现抗拷贝攻击的目的。计算机仿真结果证明了该方案的有效性和鲁棒性。

关键词：数字水印，独立分量分析，拷贝攻击
中图分类号：TP391 **文献标识码**：A **文章编号**：1009-5896(2005)07-1035-04

An ICA-Based Watermarking Scheme Resistant to Copy Attack

Hu Hui-bo Liu Ju Sun Jian-de
(*School of Information Science and Engineering, Shandong University, Jinan* 250100, *China*)

Abstract A digital watermarking scheme is proposed in this paper, which is based on Independent Component Analysis (ICA) and able to resist copy attack effectively. In this algorithm, both the image signature describing the original image uniquely and the watermark are embedded into the original image synchronously and extracted through ICA techniques. By comparing the re-calculated signature of the marked image according to the image signature algorithm with the extracted one, the legality of the marked image can be determined and the purpose of resistance to copy attack can also be realized. Simulation results demonstrate the robustness and validity of this proposed method.

Key words Digital watermarking, Independent Component Analysis (ICA), Copy attack

1 引言

近年来，数字水印技术作为保护多媒体产品产权的有效手段，已经引起越来越多的关注。用于版权保护的数字水印一般为鲁棒性水印，即嵌入水印的图像在经历各种攻击之后，仍然能够从中较好地检测或提取水印。对水印的攻击按照其目的不同，可以分为：移去攻击、几何攻击、密码攻击和协议攻击[1]。最新提出的拷贝攻击[1-3]是一种协议攻击，它并不以破坏水印或影响水印检测为目的，而是将水印从合法的水印图像中估计出来并将其复制到另外一幅没有水印的图像中去，这就给产权保护带来了新的挑战。

文献[1]给出了一种拷贝攻击的实现方法，这种攻击属于未经授权的水印嵌入问题，与密码学中发送者认证问题非常类似，但是由于嵌入图像中的信息是有效的，只是载体出现了错误，因此采用非对称密钥对水印信息进行加密并不能有效解决这一问题。一种新的思路[2]是将能够唯一描述原始图像的信息作为水印的一部分嵌入到原始图像中，但这会降低实际水印的数据有效载荷，从而导致信噪比(SNR)的减损。

本文提出的水印方案是将数字水印以及表征原始图像 ID(Identification)信息的签名同时嵌入到原始图像多层小波分解的逼近子图中，因此这不会减少代表版权信息的水印嵌入量。在水印提取过程中，本文首先采用独立分量分析(Independent Component Analysis，ICA)方法将嵌入的两种信息提取出来，从而达到检测目的。然后将提取的签名与从待测图像中计算的签名进行比较，来验证待测图像的合法性。另外，在对提取的水印及签名进行二值化处理过程中，根据它们的统计特性来设定阈值，从而进一步提高鲁棒性。计算机仿真实验表明，本文提出的方法除了对一般性的常规攻击具有较好的鲁棒性之外，更能有效地抵抗拷贝攻击。

2 图像签名和独立分量分析

2.1 图像签名的产生

作为表征原始图像 ID 信息的图像签名，描述的必须是原始图像感知方面最重要的特征，也就是说要找到签名相同但感知不同的图像必须是非常困难的[3]。

2004-01-20 收到，2004-06-14 改回
山东省优秀中青年科学家科研奖励基金(01BS04)资助课题

从图像压缩编码的角度出发，我们知道通过频域变换，原图像信号可以用其直流分量及少数低频交流分量的系数来表示。DCT 是目前图像压缩编码中应用较为广泛的一种频域变换。图像经 DCT 以后，系数大多数集中在低频分量区，而高频系数值相对来说要小的多，因此可以通过适当量化，仅对部分低频系数进行编码，从而实现图像压缩编码的目的。另外，由于人眼视觉的空间频率响应呈现低通特性，对 DCT 低频系数的较小变动，会导致图像在感知程度上发生较大变化。鉴于以上原因，我们选用原始图像 DCT 的直流和低频交流系数经二值化处理后的序列作为对原始图像的描述，即原始图像的签名[2]。

首先，将大小为 $N_1\times N_2$ 的图像分成若干个 $n_1\times n_2$ 大小的块，对每一块进行 DCT。然后按“之”字形扫描，取每块中的前 $m_1\times m_2$ 个系数。去除这 $m_1\times m_2$ 个系数中的直流分量，选剩余系数的中值作为阈值，将这 $m_1\times m_2$ 个系数二值化为 0，1 序列，作为该图像块的签名。最后再将各块的签名合并起来，可得到整幅图像的签名。我们用不同图像签名之间的汉明距离作为它们之间差异性的测度。根据随机过程和概率统计理论，不同图像对应的签名之间的平均汉明距离应为签名长度的 1/2。

我们选用 MIT 人脸库[4]中的图像来测试这种图像签名的唯一性，即要求不同图像对应的签名也应尽可能不同。本实验从人脸库中选取了 16 个人的 96 幅人脸图像，每人取 6 幅，对应头部的不同方向角度和相机镜头焦距的缩放。每一幅图像的大小为 120×128 像素，直接对其进行 DCT，取其中的前 16×16 个系数来生成签名[2]。测得不同图像签名之间的平均汉明距离为 121.34，标准离差为 11.75。可以看出应用上述算法生成的不同图像签名之间的汉明距离集中在靠近理论平均汉明距离的一个很小区域。这说明生成的签名可以唯一表征原始图像。实验结果如图 1 所示。

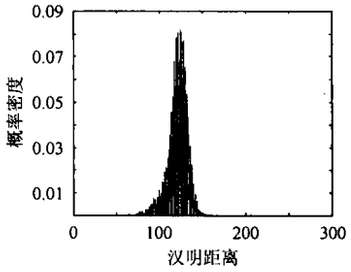

图 1 不同图像对应的图像签名之间汉明距离的分布

2.2 独立分量分析

ICA 是在研究盲源分离过程中出现的一种全新的信号处理和数据分析方法。利用 ICA 技术可以在不知道源信号和传输通道参数的情况下，根据输入源信号的统计特性，仅由观测信号恢复或提取源信号。ICA 已成为盲信号处理中最主要的方法之一，并在诸如模式识别、数据压缩、图像分析等方面得到应用[5-8]。

假设混叠系统是由 m 个传感器和 n 个源信号组成，源信号与观测信号满足：

$$\boldsymbol{x}=\boldsymbol{A}\boldsymbol{s} \tag{1}$$

其中源信号 $\boldsymbol{s}=[s_1,s_2,\cdots,s_n]^{\mathrm{T}}$ 的 n 个分量是相互统计独立的，而观测信号 $\boldsymbol{x}=[x_1,x_2,\cdots,x_m]^{\mathrm{T}}$ 是源信号 n 个分量的混叠。$m\times n$ 维的矩阵 $\boldsymbol{A}$ 为混叠权系数矩阵。

盲源分离就是求解分离矩阵 $\boldsymbol{B}$，使得通过下面的公式可以恢复源信号：

$$\boldsymbol{y}=\boldsymbol{B}\boldsymbol{x}=\hat{\boldsymbol{s}} \tag{2}$$

通常假设源信号各分量 $\{s_i,i=1,2,\cdots,n\}$ 相互统计独立。利用假设，根据测度信号独立性的准则获得分离矩阵，可以恢复出相互统计独立的源信号，这一过程因而又称为 ICA。

目前已经提出许多盲源分离算法[7]，其中基于最小互信息(Minimum Mutual Information，MMI)[8]的方法对混叠图像信号的盲分离效果较好，本文将研究基于该算法的水印提取方法。

用输出矢量各分量之间的互信息作为准则函数：

$$J(\boldsymbol{B})=\mathrm{MI}(\boldsymbol{y};y_1,y_2,\cdots,y_n)=\sum_{i=1}^{n}H(y_i)-H(\boldsymbol{y}) \tag{3}$$

由于源信号及其概率密度未知，式(3)不能直接使用。我们用 Edgeworth 展开把熵展开，并把它截短成四阶的累计量。这样处理的优点主要在于 Edgeworth 展开的各项按其对展开式的贡献递减排列，从而可以获得更好的渐近收敛性[7,8]。

基于 Edgeworth 展开的 ICA 训练公式由下式给出：

$$\boldsymbol{B}(k+1)=\boldsymbol{B}(k)-\eta(k)(\boldsymbol{I}-f(\boldsymbol{y})\boldsymbol{y}^{\mathrm{T}})\boldsymbol{B}(k) \tag{4}$$

其中 $\boldsymbol{I}$ 是单位阵，$\eta(k)$ 是步长因子，$f(\boldsymbol{y})=[f(y_1),f(y_2),\cdots,f(y_n)]^{\mathrm{T}}$ 是由 Edgeworth 展开所获得的非线性函数，其各分量可以表示为

$$f(y_i)=\left[\frac{1}{2}\mathrm{cum}_3^2(y_i^3)-\frac{1}{6}\mathrm{cum}_4(y_i^4)\right]y_i^3+\left[\frac{3}{4}\mathrm{cum}_3(y_i^3)\mathrm{cum}_4(y_i^4)-\frac{1}{2}\mathrm{cum}_3(y_i^3)-\frac{7}{4}\mathrm{cum}_3^3(y_i^3)\right]y_i^2 \tag{5}$$

式(5)中的三阶累积量 $\mathrm{cum}_3(y_i^3)$ 和四阶累积量 $\mathrm{cum}_4(y_i^4)$ 可用下面的自适应方法估计：

$$\mathrm{cum}_3(y_i^3(k+1))=\mathrm{cum}_3(y_i^3(k))+\mu(k)(y_i^3(k)-\mathrm{cum}_3(y_i^3(k))) \tag{6}$$

$$\mathrm{cum}_4(y_i^4(k+1))=\mathrm{cum}_4(y_i^4(k))+\mu(k)(y_i^4(k)-y_i^2(k)y_i^2(k-1)-\mathrm{cum}_4(y_i^4(k))) \tag{7}$$

3 水印的嵌入与提取

3.1 数字水印的嵌入

首先将原始图像进行多层小波分解，并将其逼近子图作为源信号 $\boldsymbol{s}_1$，水印和原始图像的签名分别作为源信号 $\boldsymbol{s}_2$ 和 $\boldsymbol{s}_3$。将 $\boldsymbol{s}_1$，$\boldsymbol{s}_2$ 和 $\boldsymbol{s}_3$ 表示成行向量形式，利用式(1)作为水印

嵌入模型进行混叠，其中混叠矩阵为 $\boldsymbol{A}=(a_{ij}),i=1,2,3;$ $j=1,2,3$。那么观测信号 $\boldsymbol{x}=[x_1,x_2,x_3]^{\mathrm{T}}$ 就是 s_1，s_2 和 s_3 的线性叠加，即

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{pmatrix}\begin{pmatrix}s_1\\s_2\\s_3\end{pmatrix} \tag{8}$$

如果选取 x_1 作为嵌入水印和签名信息的小波逼近子图，则 x_2 和 x_3 就是我们在提取水印和签名过程中要用到的密钥阵。对 x_1 进行小波反变换，就完成了水印的嵌入。

由于水印的嵌入在一定程度上改变了载体图像，为了避免由于水印的嵌入而导致签名无效，载体图像中要提取签名的部分不应再嵌入水印。本文之所以仍选择代表原始图像低频概貌的小波逼近子图来嵌入水印，是鉴于以下原因：(1)小波逼近子图的稳定性好，选择逼近子图来嵌入水印，可以提高水印的鲁棒性[9]。(2)由于密钥阵中含有签名信息，在检测过程中通过密钥的使用可以消除水印对签名的影响，从而使从待测图像中提取的签名与原签名保持一致。

为了满足水印的不可见性要求，应有 $a_{12},a_{13}\ll a_{11}$，同时由于密钥阵中含有图像签名信息，这使得 s_3 比 s_2 的鲁棒性要高，因此可令 $a_{12}>a_{13}$ 以进一步提高水印的鲁棒性。另外如果密钥阵 x_2 和 x_3 中含有水印信息，则在检测阶段，对于不含水印的图像，仍然能够将水印提取出来。这是由于不管源信号如何混叠，ICA 方法都可以仅通过观测信号就能恢复或提取源信号，因此为了降低虚警率，a_{22}，a_{32} 取值应为零。

为了增强保密性，我们随机产生满足上述限定条件的混叠矩阵 $\boldsymbol{A}$，其中 $(a_{12}+a_{13})\approx 0.1\cdot a_{11}$，$a_{22}=a_{32}=0$，以达到不容易被别人获取的目的。

3.2 数字水印的提取

在检测过程中，先利用式(5)训练得到分离矩阵 $\boldsymbol{B}$，再将待测图像进行多层小波分解，得到其逼近子图，然后结合密钥阵利用式(2)进行分离提取，得到水印和签名。

在这里，我们之所以不用混叠矩阵的逆矩阵和密钥阵联合起来来检测水印和签名，是因为这种检测方法对混叠矩阵的保密性不好，并且在水印检测时，由于矩阵计算问题本身有可能是病态的，微小的攻击可能会引起解的巨大变化，从而导致水印鲁棒性的降低。

由于采用 ICA 方法提取的水印和图像签名不再是 0，1 序列，故要设定阈值对其进行二值化处理。若原水印序列中 0 和 1 的元素个数比为 $k_0:k_1$，则选提取的水印中第 k_0 个最小的元素作为其二值化阈值，将大于此值的元素置为 1，其余元素置为 0。图像签名可以直接选取其中值作为阈值。

为了测量提取的水印与原水印的相似程度，我们用它们之间的归一化互相关系数(Normalized Cross-correlation，NC)作为客观评价标准：

$$\mathrm{NC}=\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\boldsymbol{W}(i,j)\cdot\boldsymbol{W}'(i,j)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\boldsymbol{W}^2(i,j)\sum_{i=1}^{M}\sum_{j=1}^{N}\boldsymbol{W}'^2(i,j)}} \tag{9}$$

其中，$\boldsymbol{W}(i,j)$，$\boldsymbol{W}'(i,j)$ 分别表示原始水印和二值化后的提取水印，水印大小为 $M\times N$。如果 NC 值大于设定的阈值 T，就认为待测图像中有水印存在；反之，则不存在水印。

4 拷贝攻击的实现与抗击

本文采用文献[1]提出的拷贝攻击的实现方法，即先从合法的水印图像 $\boldsymbol{G}_{1w}$ 中估计出原始图像 $\hat{\boldsymbol{G}}_1$，再通过从水印图像 $\boldsymbol{G}_{1w}$ 中减去原始图像的估计值 $\hat{\boldsymbol{G}}_1$ 来估计水印 $\hat{\boldsymbol{W}}$：

$$\hat{\boldsymbol{W}}=\boldsymbol{G}_{1w}-\hat{\boldsymbol{G}}_1 \tag{10}$$

最后将估计水印 $\hat{\boldsymbol{W}}$ 与未加水印的非法图像 $\boldsymbol{G}_2$ 直接相叠加，得到拷贝攻击后的水印图像 $\boldsymbol{G}_{2w}$：

$$\boldsymbol{G}_{2w}=\boldsymbol{G}_2+\hat{\boldsymbol{W}} \tag{11}$$

在本实验中，估计到的水印 $\hat{\boldsymbol{W}}$ 实际上是真实水印 s_2 与原始图像签名 s_3 的线性叠加：

$$\hat{\boldsymbol{W}}=s_2+s_3 \tag{12}$$

对于拷贝攻击的水印图像，我们仍然采用 ICA 方法进行检测。计算从待测图像中提取出来的签名与从中按签名算法得到的签名之间的汉明距离，通过与设定的阈值之间的比较来确定嵌入的水印是否属于待测图像。如果大于阈值，则说明水印是从别的图像中拷贝过来的，从而实现了抗拷贝攻击的目的。

5 计算机仿真实验结果

原始图像是 256×256 的标准“Lena”图像，如图 2(a)所示，取其三阶小波分解的逼近子图(如图 2(b))来嵌入水印。将原图像进行 128×128 分块 DCT，按“之”字形扫描，取每块的前 16×16 个系数来生成块签名，继而得到原始图像的签名，如图 2(c)所示。水印是大小为 32×32 的二值图像，如图 2(d)所示。嵌入水印的图像如图 2(e)所示。实验中，混叠矩阵 $\boldsymbol{A}=\begin{pmatrix}1&0.15&-0.02\\0.8&0&0.4\\0.06&0&1\end{pmatrix}$，水印的判决阈值取为 0.85，图像签名的判决阈值取为 470。

图 3(a)是将“Lena”水印图像与原始图像的差值信息(如图 3(b))加权后(本实验加权因子为 0.3)直接与“Cameraman”

(a) 原始图像　(b) 小波逼近子图 (c) 图像签名 (d) 水印 (e) 嵌入水印的图像

图 2

(a) 拷贝攻击后的图像 (b) 嵌入水印的“Lena”图像与原始图像的差值 (c) 提取的水印

图3 拷贝攻击后的图像及其提取的水印

图4 剪切攻击后的图像

图像相叠加得到的非法嵌入水印的图像，即为经历了拷贝攻击之后的图像。图3(c)是采用ICA方法从非法图像中提取的水印。

表1是在一般性的常规攻击下，用ICA方法和用混叠矩阵求逆的方法提取的水印与原水印之间NC值的比较，其中(a)是均值为0.1，方差为0.0015的高斯噪声攻击，(b)是3×3中值滤波，(c)是剪切掉“Lena”图像的眼睛部分，对应的攻击图像如图4所示，(d)为JEPG压缩(Gr=16.624)，(e)是采用“Bilinear”方法先将图像扩大为原图像的2倍，再缩为原图像大小。从表中可以看出，虽然在上述攻击下，ICA方法和混叠矩阵求逆法提取的水印与原水印之间的NC值都大于给定的阈值，但在相同的攻击条件下，用ICA提取的水印与用混叠矩阵的逆矩阵提取的水印相比，与原水印的相关性更强一些，从而证明采用ICA技术可以获得更好的鲁棒性。

表1 常规攻击下用不同方法提取的水印与原水印之间的NC值比较

	提取水印方法	无攻击	(a)	(b)	(c)	(d)	(e)
NC	ICA方法	1	0.9796	0.9567	0.9639	0.9423	0.9772
	混叠矩阵求逆的方法	1	0.9736	0.9507	0.9639	0.9459	0.9603

表2为上述常规攻击以及拷贝攻击下用ICA方法从攻击图像中提取的签名与从中计算的签名及原图像签名之间的汉明距离。可以看出由于水印的嵌入，即使在没有攻击的情况下，提取的签名也与从嵌入水印的图像中计算的签名有所差异，但是由于密钥阵中含有原签名的信息，从而可以保证在存在攻击的情况下，提取的签名仍然与原图像签名完全相同。另外，表2中第2行数据显示，与其他常规的攻击相比，从拷贝攻击后的图像中提取的签名与从中计算的签名之间的汉明距离值高于给定的阈值，即出现了拷贝攻击。

表2 各种攻击下提取的签名与原签名及待测图像签名之间的汉明距离

	无攻击	(a)	(b)	(c)	(d)	(e)	拷贝攻击
原签名与提取的签名	0	0	0	0	0	0	0
提取的签名与待测图像签名	16	34	28	130	34	26	542

6 结束语

本文提出了一种在不降低水印数据有效载荷的情况下，利用图像签名来抵抗拷贝攻击的水印方案。水印检测采用ICA方法，从而将盲源分离理论应用到水印领域，为水印技术的发展提供了新的思路。计算机仿真结果表明这种方法具有较好的鲁棒性。对于如何进一步设计混叠模型，提高水印的鲁棒性，以及这种方案在视频和音频作品中的应用是我们下一步的研究目标。

参 考 文 献

[1] Kutter Martin, Voloshynovskiy Sviatoslav, Herrigel Alexander. The watermark copy attack. Proc. of the SPIE, Security and Watermarking of Multimedia Contents II, San Jose, CA, 2000, 3971: 371 – 380.

[2] Barr John, Bradley Brett, Hannigan Brett T. Using digital watermarks with image signatures to mitigate the threat of the copy attack. ICASSP03, 2003 IEEE International Conference, Hong Kong, 2003, 3: 69 – 72.

[3] Cox Ingemar J, Miller Matthew L, Bloom Jeffrey A 著. 王颖, 黄志蓓译. 数字水印. 北京: 电子工业出版社, 2003: 第9章.

[4] Turk M, Pentland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1991, 3(1): 71 – 86.

[5] Yu Dan, Sattar Farook, Ma Kai-Kuang. Watermark detection and extraction using independent component analysis method. EURASIP Journal on Applied Signal Processing, 2002, 1: 92 – 104.

[6] Shen Minfen, Zhang Xinjung, Sun Lisha, Beadle P J, Chan F H Y. A method for digital image watermarking using ICA. ICA 2003, Nara, Japan, 2003: 209 – 214.

[7] Liu Ju, Nie Kaibao, He Zhenya. Blind separation by redundancy reduction in a recurrent neural network. *Chinese Journal of Electronics*, 2001, 10(3): 415 – 419.

[8] He Zhenya, Liu Ju, Yang Luxi. Blind separation of images using Edgeworth expansion based ICA algorithm. *Chinese Journal of Electronics*, 1999, 8(3): 278 – 282.

[9] 周亚训, 叶庆卫, 徐铁峰. 基于小波和余弦变换组合的图像水印方案. 电子学报, 2001, 29(12): 1693 – 1695.

胡慧博： 女，1980年生，硕士生，研究方向为多媒体通信与信息处理.

刘 琚： 男，1965年生，教授，博士生导师，通信工程系主任，山东大学—美国德州仪器(TI)公司数字信号处理实验室主任，从事盲信号处理、通信信号处理、数字水印和图像复原等研究，发表论文80余篇.

孙建德： 男，1978年生，博士生，研究方向为多媒体信号处理，发表论文10余篇.