一种解决版权问题的数字水印方案 [1]

钟 桦 刘 芳 焦李成

(西安电子科技大学雷达信号处理国家重点实验室 西安 710071)

摘 要 大部分水印技术侧重于水印的不可见性和稳健性，而版权保护问题并未得到真正解决. 该文提出的数字水印方案可以有效地解决这一问题. 通过一种新颖的水印构造方法，有效地消除了在解决版权问题中对使用原始图像的限制. 盲与非盲水印检测时水印的不可见性和稳健性均得到很大提高. 实验结果证明该数字水印方案对于版权攻击是有效的，对各种失真都具有较强的稳健性.

关键词 数字水印，版权保护，子空间分解，盲与非盲水印检测

中图号 TP391.4

1 引 言

近几年在数字水印这一领域的研究非常活跃. 根据其应用，数字水印技术可分为版权保护，数据管理和跟踪，数据鉴定等方面，其中版权保护是提出数字水印这一技术的初衷. 到目前为止已经提出了很多种水印算法并具有较好的不可见性和稳健性 [1] . 但是这些算法都忽略了解决版权问题这一最初目的.

数字水印是一个把相关信息或签名嵌入媒体数据的过程，其中嵌入信息可以是版权信息，标识或辅助说明. 一般认为只要从媒体数据中检测到或提取出所嵌入的水印，即可确定其合法拥有者. 但是早在文献 [2] 中就已经明确指出，仅依靠稳健的水印算法并不能解决版权问题. 文献 [3] 中构造了几种攻击方法可以使大部分存在的水印算法在解决版权问题上失效，并进一步提出了一种准不可逆水印算法框架. 作者称该水印框架可以解决版权问题，但是其缺点是过于复杂，且并没有给出相应的证明. 文献 [4] 认为版权问题的解决关键在于不能使用原始的媒体数据，并提出了一种优化的盲水印检测技术.

以图像为例，本文提出了一种新颖的数字水印方案. 所构造的水印可以有效地解决版权问题并彻底消除盲与非盲水印检测之间的差别. 实验结果证明所提出的水印方案对各种失真以及版权攻击是稳健的.

2 水印与版权问题

2.1 水印算法分析

给定原始图像 I 和水印 S，由水印嵌入算法可以得到加水印图像 $\hat{I}$. 在水印检测过程中通过计算提取的水印 W 和原始水印 S 的相似程度并与设定的门限比较，以判断水印的存在与否，从而确定该图像的版权归属. 这似乎可以用来解决版权问题，但事实并非如此.

例 1[3] 假设嵌入过程是可逆的，则攻击者可以对加水印图像 $\hat{I}$ 逆向嵌入水印 S'，得到伪造的原始图像 $\hat{I}'$，即出现死锁 (deadlock) 问题 [5] .

例 2[4] 攻击者宣称加水印图像 $\hat{I}$ 是其原始图像，而原始图像 I 是由 $\hat{I}$ 逆向嵌入水印而得到的. 因此版权问题仍无法得到解决.

对于例 1，文献 [3] 中指出嵌入过程的可逆特性是其根本原因，并进一步提出嵌入过程必须是准不可逆的才能彻底解决版权问题. 这样例 1 和例 2 中的情况都可以得到解决. 但是对水印嵌入过程的准不可逆性的证明非常复杂，文献 [3] 中也没有给出严格的证明. 文献 [4] 则认为版

[1] 2001-08-14 收到，2002-04-15 改回
国家自然科学基金项目 (60073053)

权问题的本质在于水印检测过程不能使用原始图像，其依据是原始图像 I 没有经过任何处理，因此不可能检测到任何水印，从而可以证明原始图像的真实性．设原始水印为 S_1，$\{I_i\}$ 是由原始图像 I 中提取出的特征集，水印嵌入可以表示如下：

$$I_i' = I_i + G_i(I_i)S_{1i} \tag{1}$$

其中 $G_i(I_i)$ 是水印的能量加权．水印检测过程利用随机序列 S_2，根据假设检验判断水印存在与否．

$$\begin{aligned} H_0 &: X_i = I_i + N_i, && \text{无水印} \\ H_1 &: X_i = I_i + G_i(I_i)S_{1i} + N_i, && \text{有水印} \end{aligned} \tag{2}$$

其中 N_i 为噪声，相关检测器输出为

$$q = M_y\sqrt{n}/V_y \tag{3}$$

其中 $y = \{Y_i\}$, $Y_i = X_iS_{2i}$, n 是特征集 $\{I_i\}$ 的长度．M_y 和 V_y 分别定义如下：

$$M_y = \sum_{i=1}^{n} Y_i \Big/ n, \quad V_y^2 = \sum_{i=1}^{n}(Y_i - M_y)^2 \Big/ (n-1) \tag{4}$$

随机序列 S_2 必须满足：(1) 均值为 0; (2) 与原始图像 I 不相关；(3) 与 S_1 具有很强的相关性．由此可得在假设 H_0 时，q 服从标准正态分布，即 $q \sim N(0,1)$; 而在假设 H_1 时，q 服从正态分布 $N(m,1)$．通过优化设计 S_2 可以使 q 具有较大的均值 m．为避免攻击者通过穷举法找到一个能产生较大 q 值的随机序列作为假水印，文献 [4] 对序列 S_2 再附加两个条件：(4) 必须使用注册的标识符 (ID) 或有意义的签名 (signature) 作为 S_1 的种子；(5) 利用单向函数把种子映射为 S_1(再由 S_1 得到 S_s)．可以看出，水印的稳健性将取决于条件 (1)—(3)，而条件 (4)，(5) 的目的是限制水印的选取范围．

2.2 存在的问题

通过分析可以提出以下两种情形：

例 3　条件 (4), (5) 可能与条件 (2) 相矛盾，使得产生的水印与原始图像 I 具有较强的相关性，即出现虚警．

例 4　攻击者由加水印图像 $\hat{I}$ 中减去一个随机序列 S'，即 $\hat{I}_i' = \hat{I}_i - S'$．以 $\hat{I}'$ 作为伪造的原始图像，按照合法的水印算法嵌入标准水印 S'' 得到另一图像 $\hat{I}''$．S'' 满足条件 (1)—(5)，且与 S' 具有较强的相关性．

可见虽然条件 (4), (5) 在一定程度上对水印进行了限制，但同时可能与条件 (2) 相矛盾，即出现例 3 的情况．一种方法是扩大水印的选取范围，例如可以结合条件 (4) 和用另一密钥 K，随机改变密钥 K 直到产生符合要求的水印．但是这同样有利于攻击者，因为密钥 K 的随机性使得攻击者在选取水印时具有更大的自由度．

在例 4 中，假设合法所有者和攻击者的水印均符合要求且二者都能从测试图像 $\hat{I}''$ 中检测到水印，法官将要求对各自的原始图像 I 和 $\hat{I}'$ 进行鉴别．这是解决版权问题的关键．若水印算法足够稳健，合法所有者应当可以从图像 $\hat{I}'$ 中检测到水印．而攻击者从图像 I 中检测到水印的概率即等于 S'' 与 S' 可能相关的概率，与算法本身无关．可知利用水印 S'' 得到的检测器输出 q 同样满足标准正态分布，此时 q 大于检测门限的概率即为攻击概率．

3 一种新颖的水印方案

3.1 水印产生

设原始图像的特征集为 $\{I_i\}$, $\{I_i\}$ 已经过扰乱处理以确保安全性 [6] 。 $\{I_i\}$ 可写作矢量形式:

$$\boldsymbol{I}' = [I_1, I_2, \cdots, I_n]^T = [\boldsymbol{I}_1^T, \boldsymbol{I}_2^T, \cdots, \boldsymbol{I}_M^T]^T \tag{5}$$

其中 $\boldsymbol{I}_j$, $j = 1, \cdots, M$ 是长度均为 n/M 的子矢量。根据特征子空间的不变性分析可知，数据矩阵越大，特征值对数据矩阵的微小扰动越不敏感 [7] 。由于水印必须是基于原始图像的，为增大水印对原始图像的敏感性， $\boldsymbol{I}_j$ 的长度不能太大。对 $\boldsymbol{I}_j$ 的协方差矩阵 $\boldsymbol{R}_j$ 进行特征分解:

$$\boldsymbol{R}_j = \boldsymbol{P}_j \boldsymbol{\Lambda}_j \boldsymbol{P}_j^H \tag{6}$$

其中 $\boldsymbol{P}_j$ 是归一化特征矢量所构成的正交矩阵， $\boldsymbol{\Lambda}_j = \mathrm{diag}(\lambda_j(1), \lambda_j(2), \cdots, \lambda_j(n/M))$ 是特征值的对角矩阵。对于图像信号，可以假设 $\lambda_j(1) > \cdots > \lambda_j(K) > \lambda_j(K+1) = \cdots = \lambda_j(n/M) = 0$, $\boldsymbol{I}_j$ 可以写成

$$\boldsymbol{I}_j = \sum_{k=1}^{K} \sqrt{\lambda_j(k)} \boldsymbol{P}_j(k) \tag{7}$$

其中 $\boldsymbol{P}_j(k), k = 1, \cdots, K$ 称为信号特征矢量，记作 $\boldsymbol{P}_j^s$ ，而 $\boldsymbol{P}_j(k)$, $k = K+1, \cdots, n/M$ 称为噪声特征矢量，记作 $\boldsymbol{P}_j^n$ 。根据正交矩阵的特性， $\boldsymbol{P}_j^n$ 与 $\boldsymbol{P}_j^s$ 完全正交，因此

$$\boldsymbol{I}_j^T \cdot \boldsymbol{P}_j^n(k) = 0, \quad k = K+1, \cdots, n/M \tag{8}$$

任取一组整数 k_j, $K+1 \le k_j \le n/M$ 。令 $\boldsymbol{S}_1 = [\boldsymbol{P}_j^n(k_1)^T, \cdots, \boldsymbol{P}_M^n(k_M)^T]^T$ ，可以得到

$$\boldsymbol{I}'^T \cdot \boldsymbol{S}_1 = \sum_{j=1}^{M} \boldsymbol{I}_j^T \cdot \boldsymbol{P}_j^n(k_j)^T = 0 \tag{9}$$

(9) 式表明 $\boldsymbol{S}_1$ 是基于原始图像的且与主数据干扰完全正交，其产生过程也具有一定程度的不可逆性，因此能够同时满足条件 (1)–(5) 。由于 $\boldsymbol{S}_1$ 由归一化的噪声特征矢量构成， $\boldsymbol{S}_1$ 的均值约为 0 ，且方差为

$$E(\boldsymbol{S}_{1i}^2) = \frac{1}{n} \sum \boldsymbol{S}_{1i}^2 = \frac{1}{n} \sum_{j=1}^{M} \boldsymbol{P}_j^n(k_j)^T \boldsymbol{P}_j^n(k_j) = \frac{M}{n} \tag{10}$$

为保证水印本身的能量足够大，令 $S_1 = \sqrt{n/M}\boldsymbol{S}_1$, 以 S_1 作为水印，其方差为 1 。

3.2 水印嵌入和检测

水印按照 (1) 式进行嵌入。水印检测分为盲与非盲水印检测技术两种，以下只考虑盲水印检测算法。令 $S_2 = \boldsymbol{S}_1$ ，在假设 H_0 时，在版权问题中，合法所有者和攻击者对各自的原始图像进行水印检测，因此可以不考虑噪声 N_i 。由于 $\boldsymbol{S}_1$ 与特征集矢量 $\boldsymbol{I}'$ 的正交性，可得

$$M_y = \sum_{i=1}^{n} Y_i \Big/ n = (1/n)\boldsymbol{I}'^T \cdot S_1 = 0 \tag{11}$$

此时 $q=0$．在假设 H_1 时，$Y_i=X_iS_{2i}=I_iS_{2i}+G_i(I_i)S_{1i}S_{2i}+N_i$，不考虑噪声 N_i，并假设 $\boldsymbol{S}_{1i}^2$ 与 $G_i(I_i)$ 独立，可以得到

$$M_y=\frac{1}{n}\sum_{i=1}^{n}(I_iS_{2i}+G_i(I_i)S_{1i}S_{2i})\approx\sqrt{\frac{n}{M}}E(G_i(I_i))E(\boldsymbol{S}_{1i}^2) \tag{12}$$

$$V_y^2\approx E(I_i^2\boldsymbol{S}_{1i}^2) \tag{13}$$

可见均值 M_y 完全不受原始图像的影响，而方差 V_y^2 主要与原始图像和水印有关．可以假设 I_i^2 也与 $\boldsymbol{S}_{1i}^2$ 独立．q 的均值 m 可由下式得到

$$m\approx E(G_i(I_i))\sqrt{n}/\sqrt{E(I_i^2)} \tag{14}$$

3.3 性能分析

下面从虚警概率和攻击概率两个方面分析所提出水印方案．虚警概率是指在没有水印的情况下却检测到水印的概率．由于水印只能由原始图像产生并加载于该图像，由 (11) 式可知对于合法所有者不存在虚警现象，即虚警概率为 0．

根据例 4，攻击概率与水印嵌入算法无关．攻击者首先从加水印图像 $\hat{I}$ 中减去加权的随机序列 $G_i(I_i)S_i'$，其中 $S'\sim N(0,1)$．然后按本文方法由伪造的原始图像 $\hat{I}'$ 产生标准的水印 S''．把 $G_i(I_i)S_i'$ 看作噪声 N_i，可知在非盲水印检测中输出 q 将服从均值为 0，方差为 1 的正态分布 [2,4]．设盲与非盲水印检测中由 (4) 式计算的方差分别为 V_y^2 和 $\tilde{V}_y^2$，根据 (4) 式可得

$$V_y^2\approx E(Y_i^2)-M_y=E((I_i+N_i)^2S_{2i}^2)-M_y^2=\tilde{V}_y^2+E(I_i^2S_{2i}^2)+2E(I_iN_iS_{2i}^2) \tag{15}$$

其中 $\tilde{V}_y^2\approx E(N_i^2S_{2i}^2)-M_y^2$．因为 I_i 与 S_{2i} 独立，可以假设 I_i^2 与 S_{2i}^2 独立，则上式中的第二项和第三项可以写成：

$$E(I_i^2S_{2i}^2)+2E(I_iN_iS_{2i}^2)=E(I_i^2)E(S_{2i}^2)+2E(I_i)E(N_i)E(S_{2i}^2)=E(I_i^2)E(S_{2i}^2) \tag{16}$$

所以

$$V_y^2/\tilde{V}_y^2\approx 1+E(I_i^2)E(S_{2i}^2)/(E(N_i^2S_{2i}^2)-M_y^2) \tag{17}$$

因为 N_i 与 S_{2i} 独立，同样可以假设 N_i^2 与 S_{2i}^2 独立．注意到 $M_y\approx 0$ 且 $E(I_i^2)$ 远大于 $E(N_i^2)$，则 (17) 式可以写成

$$\tilde{V}_y^2\approx V_y^2[E(N_i^2)/E(I_i^2)] \tag{18}$$

令 $\beta=E(N_i^2)/E(I_i^2)$．由于 $N_i=G_i(I_i)S_i'$，$S_i'\sim N(0,1)$，可以假设 S_i' 与 $G_i(I_i)$ 独立以及 $S_i'^2$ 与 $G_i^2(I_i)$ 独立，则 β 可以估计如下：

$$\beta=E(G_i^2(I_i)S_i'^2)/E(\hat{I}_i'^2)\approx E(G_i^2(I_i))/E(I_i^2) \tag{19}$$

假设最大能量加权取 $G_i(I_i)=\alpha|I_i|$，则 (19) 式可以简化为

$$\beta\approx\alpha^2E(I_i^2)/E(I_i^2)=\alpha^2 \tag{20}$$

显然 $\beta<1$．对 (3) 式进行如下修正：

$$q'=M_y\sqrt{n}/(V_y\sqrt{\beta}) \tag{21}$$

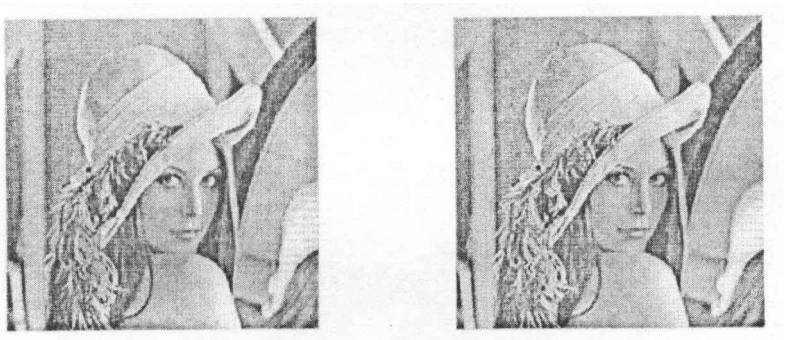
此时检测器输出 q' 服从标准正态分布。在这里 β 起方差的放大作用而并不影响真实值。q' 大于检测门限 T 的概率即为攻击概率。门限 T 可根据不同应用来选择。由于攻击者的水印至少是二次水印，其能量加权不能取最大以保证图像的感知质量。因此对参数 β 的估计值偏大，导致 q' 的方差小于 1 。这意味着攻击概率降低。由此可见，提高水印的加权能量是水印算法的一个关键问题。为统一起见，本文实验中恒取 $\beta=\alpha^2$ ，则在假设 H_1 时检测器输出的均值为

$$m' \approx E(|I_i|)\sqrt{n}/\sqrt{E(I_i^2)} \tag{22}$$

可见均值 m' 只与特征集及其长度 n 有关。n 越大则 m' 越大。在 $G_i(I_i)=\alpha|I_i|$ 时，要求特征集 $\{I_i\}$ 幅度绝对值的均值尽量大而方差尽量小。

4 实验结果

实验中原始图像为 $256\times256\times8$ 的标准 Lena 图像。取其小波分解的第三层所有小波系数作为特征集，长度 $n=3072$ 。经伪随机扰乱后按照 (1) 式进行修改，其中取 $M=48,\alpha=0.1$ 。取 $G_i(I_i)=\alpha|I_i|$ ，参数 $\beta=\alpha^2$ 。嵌入结果如图 1 所示，其中加水印图像的峰值信噪比 (PSNR) 高达 51.15dB，具有很好的感知质量。



(a) 原始图像　　(b) 加水印图像 (PSNR=51.15dB)

图 1

4.1 算法稳健性

分别对加水印图像进行盲与非盲水印检测。由于水印 S_1 的特性 (与特征集 $\{I_i\}$ 完全正交)，限制使用原始图像已无意义。盲水印检测具有更大的实用性，如用于实时检测或公开密钥水印检测。非盲水印检测中利用原始图像可以使水印更加稳健。图 2 中分别给出在两种检测方法中水印对 JPEG 压缩的稳健性。可见输出值均大于文献 [4] 中相同尺寸甚至大 2 倍的 Lena 图像得到的结果，其中当质量因子 $Q=10$ 时 q 仍高达 27.42 。表 1 给出了在无失真，缩放 (缩小 1/2 以及放大 2 倍和 4 倍) 和中值滤波 (3×3) 时的稳健性测试结果，其中最小输出为 $q=20.64$ 。这说明本文所提出的水印方案是稳健的。

表 1　水印检测结果

失真类型	无失真	图像缩放			中值滤波 (3×3)
		缩小 1/2	放大 2 倍	放大 4 倍	
盲水印检测	116.00	20.64	53.61	30.82	58.00
非盲水印检测	158.06	21.09	57.95	31.83	65.47

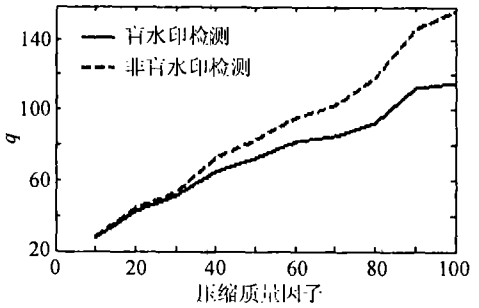


图 2　在 JPEG 压缩失真时盲与非盲水印检测器输出

4.2 版权攻击下的算法稳健性

在没有受到类似于例 4 版权攻击的情况下，合法所有者只需通过盲水印检测算法对存在质疑的待测图像进行水印检测即可.

假设根据例 4 进行版权攻击并重复 10000 次. 每次取不同的随机序列 S', $S' \sim N(0,1)$. 相应可以产生 10000 幅伪造的原始图像 $\hat{I}'$. 如果无法从感知质量上辨别真假原始图像 I 和 $\hat{I}'$, 法官要求合法所有者和攻击者分别从对方的原始图像中检测水印. 图 3 给出了检测结果，其中只给出了攻击者对加水印图像 $\hat{I}$ 的盲与非盲水印检测结果. 如果攻击者不能从 $\hat{I}$ 中检测到水印，就更加不可能从原始图像 I 中检测到水印. 由于盲与非盲水印检测的输出具有相同的分布，在图 3 中互相叠加，因此只以 $q1$ 表示. 可见 $q1$ 近似服从标准正态分布，这证明 3.3 节中的推导是正确的. 合法所有者对不同图像 $\hat{I}'$ 的盲与非盲水印检测的输出分别为 $q2$ 和 $q3$, 其均值分别为 115.40 和 156.58 , 并近似满足方差为 1 的正态分布.

显然攻击者要在合法所有者的原始图像中 "嵌入" 假水印几乎是不可能的. 与文献 [4] 中的结果相比较 (盲水印检测时，在无失真的条件下输出均值为 28.4) , 在方差不变的情况下，合法所有者和攻击者的检测器输出均值差距更大. 由于实验中基于最简单的能量加权模型，特征集长度 (3072) 也要远小于文献 [4] 中的特征集长度 (62338) , 本文水印方案的优点显而易见.

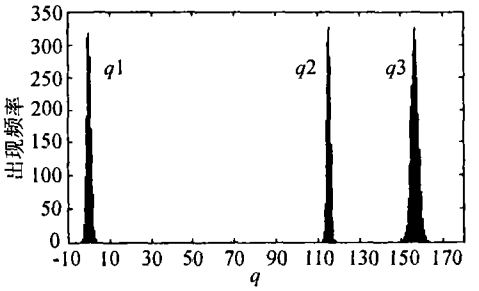


图 3　攻击下的检测器输出分布

$q1$: 攻击者的盲与非盲检测器输出;　$q2$: 合法所有者盲水印检测器输出;　$q3$: 合法所有者非盲水印检测器输出

5 结　论

本文提出了一种有效的用于解决版权问题的数字水印方案. 利用新颖的构造方法产生的水印可以有效地解决版权问题. 这种水印方案可以消除在解决版权问题时对使用原始图像的限制，

使得盲与非盲水印检测技术都可用于解决版权问题并具有不同的应用。所提出的数字水印方案较之文献 [3, 4] 更加简单有效。虽然所使用的水印能量加权基于很简单的视觉模型，但是在水印的不可见性和稳健性上都是令人满意的。本文还构造了一种可能的攻击方案以估计攻击概率，从而评估水印方案的性能。实验结果证明该数字水印方案对于解决版权问题是有效的。

参 考 文 献

[1] F. Hartung, M. Kutter, Multimedia watermarking techniques, Proc. IEEE, 1999, 87(7), 1079–1107.

[2] I. J. Cox, J. Kilian, F. T. Leighton, T. Shamoon, Secure spread spectrum watermarking for multimedia, IEEE Trans. on Image Processing, 1997, IP-6(12), 1674–1687.

[3] S. Craver, N. Memon B. Yeo, M. M. Yeung, Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks and implications, IEEE J. on Selected Areas in Communications, 1998, 16(4), 573–586.

[4] W. Zeng, B. Liu, A statistical watermark detection technique without using original image for resolving rightful ownerships of digital images, IEEE Trans. on Image Processing, 1999, IP-8(11), 1534–1548.

[5] M. D. Swanson, B. Zhu, A. H. Tewfik, Multiresolution scene-based video watermarking using perceptual models, IEEE J. on Selected Areas in Communications, 1998, 16(4), 540–550.

[6] Chiou-Ting Hsu, Ja-Ling Wu, Hiding digital watermarks in image, IEEE Trans. on Image Processing, 1999, IP-8(1), 58–68.

[7] F. S. V. Bazan, Sensitivity eigenanalysis for single shift-invariant subspace-based method, Signal Processing, 2000, 80(1), 89–100.

A WATERMARKING SCHEME FOR RESOLVING RIGHTFUL COPYRIGHT

Zhong Hua Liu Fang Jiao Licheng

(*National Key Lab. for Radar Signal Processing, Xidian University, Xi'an* 710071, *China*)

Abstract Most watermarking schemes focus on the robustness and invisibility with the problem of copyright protection remained unresolved, which are their initial motivation. In this paper, a watermarking scheme is proposed to provide effective copyright protection. By using a novel method of generating watermark, the use of original data is no longer limited in resolving rightful ownerships. Moreover, blind and non-blind watermark detections have different applications. The robustness and invisibility are also improved. Both theoretical analysis and experimental results show its effectiveness against copyright attack and other distortions such as JPEG compression, Additive White Gaussian Noise (AWGN), image resizing and media filtering.

Key words Digital watermarking, Copyright protection, Subspace decomposition, Blind and non-blind watermark detection

钟 桦： 男，1976 年生，博士生，主要研究领域为智能信息处理，信息隐藏，数字水印.
刘 芳： 女，1963 年生，副教授，主要研究领域为智能信息处理，模式识别，电子商务，数字水印.
焦李成： 男，1959 年生，教授，博士生导师，主要研究领域为智能信息处理，非线性理论，数字水印.