

文章编号:1001-9081(2010)02-0517-04

一类基于奇异值分解的图像水印算法伪验证分析

赵星阳,孙继银

(第二炮兵工程学院 402 室, 西安 710025)

(sharpsword@126.com)

摘要:通过分析发现,一类基于奇异值分解的图像水印算法存在致命漏洞,即利用在水印嵌入过程中生成的密钥可以从其他图像(含未嵌入水印的图像)中提取出高相关的水印信息,称这一新发现的算法漏洞为水印算法的伪验证,并由此首次提出了水印算法伪验证的定义和判定条件。然后通过实例分析,指出基于奇异值分解的水印算法之所以存在伪验证,是由于奇异值分解使得水印信息主要包含于正交阵 U 、 V 中,因此水印的提取与正交阵 U 、 V 以及奇异值的分布类型有关,而与奇异值的具体取值无关。

关键词: 奇异值分解; 图像水印; 伪验证

中图分类号: TP309; TP391.41 文献标志码:A

Fake verification analysis of SVD-based image watermarking

ZHAO Xing-yang, SUN Ji-yin

(402 Staff Room, The Second Artillery Engineering College, Xi'an Shaanxi 710025, China)

Abstract: A fatal bug of a kind of Singular Value Decomposition (SVD)-based image watermarking was discovered in this paper. That is, using the same secret key which is produced in the embedding procedure, highly correlative watermark information can be extracted from other different images, even including non-watermarked images. This newly discovered bug was named as fake verification of watermark algorithm. At the same time, the definition and judging condition of fake verification were also presented. Then, the experimental results reveal that the watermark information is involved in the orthogonal matrix U , V after SVD. And the extracted watermark information is mainly involved in U , V , and the distribution of single value instead of exact single value. That is the reason why the SVD-based watermark algorithm has the bug of fake verification.

Key words: Single Value Decomposition (SVD); image watermarking; fake verification

0 引言

Liu 和 Tan^[1-2]最早提出基于奇异值分解(Singular Value Decomposition, SVD)的数字图像水印算法,其基本思想是:利用奇异值的微小扰动来嵌入水印信息。在此基础上,陆续出现了基于奇异值分解的新水印算法。有在 Liu 和 Tan 基础上改进的空域奇异值分解水印算法^[3-4],有基于宿主图像离散余弦变换(Discrete Cosine Transform, DCT)系数奇异值分解的水印算法^[5-7],有基于宿主图像离散小波变换(Discrete Wavelet Transform, DWT)系数奇异值分解的水印算法^[8]。从上述文献给出的仿真结果来看,这些基于奇异值分解的水印算法对压缩、旋转、剪切等操作具有很好的鲁棒性,看起来似乎是一类很好的水印算法。

但通过分析发现,上述文献仅仅是对进行了压缩、旋转等仿真操作后的图像水印作品进行检测,并没有考察利用相同的密钥从其他图像作品中提取水印的情形,实际上水印测试工具 StirMark^[9-10]也没有考察此类情形。然而经过试验发现,此类情形恰好暴露出上述文献中的水印算法存在致命的漏洞,即利用上述算法在水印嵌入过程生成的密钥可以从其他任意的图像(包括未嵌入任何水印的图像)中提取出高相关的水印信息(有意义的水印图像标识或随机水印信号),换言之,水印信息的提取和密钥高度相关,而和图像作品关联较

小。这显然不符合水印提取时,水印信息只能从确实嵌入了水印的作品中检测到或提取出的实际应用要求。因此上述文献中的水印算法对压缩、旋转、剪切等操作表现出来的强鲁棒性不具有真实性。

由此,引出水印算法设计中一个新出现的重要问题:水印的伪验证,即表面上看起来没有问题的水印算法是否存在设计上的漏洞,使得水印信息并不是从水印作品中真实提取出的,从而造成水印信息的伪验证。本文首先描述基于奇异值分解的水印算法的具体内容以及存在的漏洞,然后给出水印伪验证的定义和存在的判定条件,最后对上述水印算法进行分析,指出基于奇异值分解时这些水印算法存在伪验证的根本原因,并给出相关例证。

1 基于奇异值分解的图像水印算法漏洞

由于对图像数据进行 SVD 时,得到的奇异值表示的是亮度,对奇异值作更改时,图像亮度将发生相应变化,因此对亮度信号进行调整看起来是一种值得采用的水印嵌入方式。本章所描述的水印算法^[1-8]正是基于这一思想,均采用了奇异值微小扰动的水印嵌入方式。根据水印信息与宿主图像奇异值融合方式(扰动方式)上的不同,目前基于奇异值分解的图像水印算法可以分为三类:

1)宿主图像经过奇异值分解后,水印信息按照加性方式

收稿日期:2009-08-24。

作者简介:赵星阳(1980-),男,四川南充人,博士研究生,主要研究方向:数字水印; 孙继银(1952-),男,山东单县人,教授,博士生导师,CCF 高级会员,主要研究方向:计算机检测与控制、计算机仿真、图像处理、虚拟现实。

叠加在其奇异值矩阵上。即：

$$\begin{cases} I \Rightarrow USV^T \\ S + \alpha w \Rightarrow U_w S_w V_w^T \\ I^* \Leftarrow US_w V^T \end{cases} \quad (1)$$

注意,在实际运用中,进行奇异值分解前还可能会对宿主图像或水印信息施加 DCT 或 DWT,此处忽略此类变换的描述,主要对水印算法如何在奇异值上叠加水印信息的关键过程进行描述,下文同此。令 $K_e = \{U_w, V_w, S\}$, K_e 即为嵌入过程中生成的密钥,需要在水印提取时使用。有:

$$\begin{cases} I^* \Rightarrow U \hat{S}_w V^T \\ \hat{w} \Leftarrow \frac{1}{\alpha} (U_w \hat{S}_w V_w^T - S) \end{cases} \quad (2)$$

文献[1~2,6]中的水印算法属于此类。

2) 宿主图像和水印信息都经过奇异值分解后,水印信息的奇异值按照加性或乘性方式叠加在宿主图像的奇异值上。其中,加性叠加方式下,有:

$$\begin{cases} I \Rightarrow USV^T \\ w \Rightarrow U_w S_w V_w^T \\ S + \alpha S_w \Rightarrow S_1 \\ I^* \Leftarrow US_1 V \end{cases} \quad (3)$$

令 $K_e = \{U_w, V_w, S\}$ 为生成的密钥。水印提取时,有:

$$\begin{cases} I^* \Rightarrow U \hat{S} V^T \\ \hat{w} \Leftarrow \frac{1}{\alpha} U_w (\hat{S} - S) V_w^T \end{cases} \quad (4)$$

文献[3,7~8]中的水印算法属于此类。

而在乘性叠加方式下,有:

$$\begin{cases} I \Rightarrow USV^T \\ w \Rightarrow U_w S_w V_w^T \\ S(1 + \alpha S_w) \Rightarrow S_1 \\ I^* \Leftarrow US_1 V \end{cases} \quad (5)$$

令 $K_e = \{U_w, V_w, S\}$ 为生成的密钥。水印提取时,有:

$$\begin{cases} I^* \Rightarrow U \hat{S} V^T \\ \hat{w} \Leftarrow \frac{1}{\alpha} U_w (\hat{S}/S - 1) V_w^T \end{cases} \quad (6)$$

文献[4]中的水印算法属于此类。

3) 宿主图像经过奇异值分解后,与水印信息根据预定义的混合矩阵进行融合。

$$\begin{cases} I \Rightarrow USV^T \\ \begin{bmatrix} Sw_1 \\ Sw_2 \end{bmatrix} \Leftarrow \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} S \\ w \end{bmatrix} \\ Sw_1 \Rightarrow U_w S_w V_w^T \\ I^* \Leftarrow US_w V^T \end{cases} \quad (7)$$

其中: $\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}$ 为满秩矩阵,用于控制水印信息和宿主

图像奇异值混合时的比例系数。令 $K_e = \{U_w, V_w, Sw_2\}$ 为生成的密钥。水印提取时,有

$$\begin{cases} I^* \Rightarrow U \hat{S}_w V^T \\ Sw_1 \Leftarrow U_w \hat{S}_w V_w^T \\ [\hat{S} \quad \hat{w}] \Leftarrow ICA(Sw_1, Sw_2) \end{cases} \quad (8)$$

其中: $ICA(\cdot)$ 为主成分提取函数。由于 \hat{S} 和 \hat{w} 都包含了水印信息 w ,因此主成分提取函数可有效地从 Sw_1 以及 Sw_2 中提

取出水印信息 \hat{w} 。文献[5]中的水印算法属于此类。

从文献[1~8]中给出的仿真结果来看,其依存的事实似乎是遭受了失真攻击的水印作品在进行 SVD 时,其奇异值未发生显著变化,因此水印信息才能被有效提取。但实际情形并非如此。图 1 给出 256×256 大小的 Lena 图像分别经过旋转 15° ,纵向 $1/2$ 剪切以及旋转 30° 并剪切后的奇异值变化对比结果。

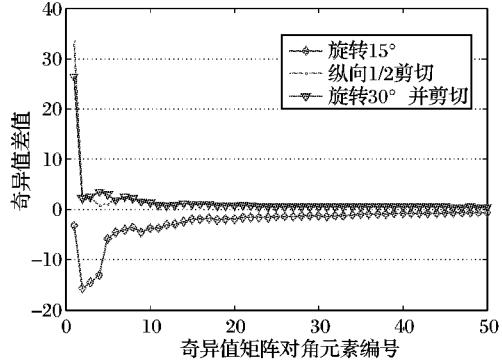


图 1 Lena 图像经旋转或剪切后的奇异值变化(前 50 个最大值)

根据文献[1~2]给出的分析证明,在微小扰动下,水印信息施加到奇异值上的扰动量约为 $\alpha \|w\|_2$,因此若水印信息的确是从奇异值的微小变化中提取出来的,那么经过旋转、剪切等操作后,奇异值的变化应近似服从方差很小的高斯分布,但从图 1 可以看出,奇异值发生了显著变化,因此这初步说明了并不是奇异值在显著影响水印信息的提取。

为进一步说明问题,利用不同水印算法在 Lena 图像中嵌入水印(水印为 Boat 图像)时生成的密钥,从未嵌入水印的 Baboon 图像中提取水印信息,结果见图 2。其中,算法 1 来自文献[1],算法 2 来自文献[3],算法 3 来自文献[5]。图 2 给出的提取结果表明,这些算法的确存在致命的漏洞。

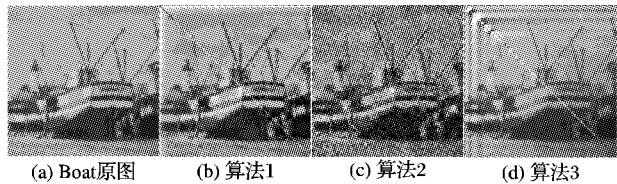


图 2 三种不同算法提取的水印信息

2 水印算法的伪验证定义

由于前文所论及的水印算法漏洞,是在考察相同的密钥从不同的作品中提取水印时暴露出来的,因此将水印算法的伪验证定义如下:

定义 1 设原始作品为 I ,水印信息为 w ,嵌入密钥为 K_u ,嵌入函数为 E ,在水印嵌入过程中会生成与 I 以及 w 相关的附加信息,将该信息定义为提取密钥 K_e ,最终生成的水印作品为 I^* ,即:

$$E: \{I, w, K_u\} \rightarrow \{I^*, K_e\} \quad (9)$$

水印作品 I^* 中包含的水印信息 \hat{w} 可通过密钥 K_u, K_e 提取,即:

$$E^{-1}: \{I^*, K_e, K_u\} \rightarrow \hat{w} \quad (10)$$

设 $\Theta(I^*)$ 为水印作品 I^* 能检测出水印的邻域空间,若 $\exists \tilde{I} \notin \Theta(I^*)$,代入式(10)后,对提取出对应的水印信息 \tilde{w} 作归一化相关性检测,若满足

$$\rho = \overline{\text{Corr}(w, \tilde{w})} = \frac{w \cdot \tilde{w}}{\sqrt{w \cdot \tilde{w}}} \geq 0 \quad (11)$$

则称该水印算法存在伪验证。而 ρ 值与密钥 K_w 中包含的水印信息量以及水印提取方式有关。

对文献[1~8]中提出的三类基于奇异值分解的水印算法,第1章已经找到存在这样的 $\tilde{I} \notin \Theta(I^*)$,且满足式(11)的条件,因此都存在伪验证的问题。不过,从定义1给出的描述来看,并不是所有的水印算法都存在伪验证。只有在嵌入过程中会生成和水印信息有关的附加信息,而且附加信息作为密钥必须在水印提取时使用的情况下,才有可能存在伪验证的问题。

3 基于奇异值分解的图像水印算法伪验证分析

从上述三类基于奇异值分解的水印算法的描述可知,虽然宿主图像和水印信息在融合方式上不尽相同,但都包含对水印信息 w 的SVD,而且分解后得到的 U_w 和 V_w 分量都被作为密钥 K_w 存储,并在提取水印信息时使用。本节将深入分析在保持 U_w 、 V_w 分量不变时,在不同的奇异值取值分布下,提取的水印信息及其与原水印信息的相关性。

分析时,令 w 为水印信息,实际应用中 w 可能为有意义的水印图像,或无意义的随机水印信号(本文将随机水印信号视作随机水印图像)。对 w 做SVD,有

$$w = U_w S_w V_w^T \quad (12)$$

令 \hat{w} 为保持 U_w 、 V_w 分量不变,奇异值发生变化时反解出的水印信息,即

$$\hat{w} = U_w S^* V_w^T \quad (13)$$

其中 S^* 为与 S_w 等大小的对角阵。分别考查 S^* 对角元素取值为全1、服从负指数分布以及服从负斜率分布三种情形下(下文将说明选择这三种取值情形的具体原因),由式(13)反解出的 \hat{w} ,并检验 w 和 \hat{w} 的相关性,即 $\rho = \text{Corr}(w, \hat{w})$, ρ 为归一化相关系数(NC)。本文用于分析的水印信息包括从文献[1~8]中选择的4个有意义水印图像标识(图3),以及100幅 64×64 大小的随机水印图像(图像数据服从 $N(0,1)$ 高斯分布)。

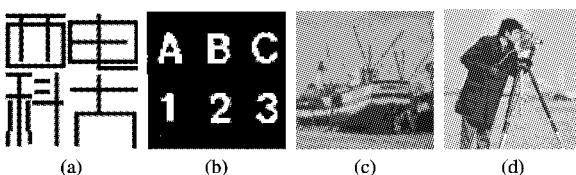


图3 用于伪验证分析的4个有意义水印图像标识

1)令 S^* 对角数据元素的取值为全1,考察 \hat{w} 的取值情况,以及检验 w 和 \hat{w} 的相关性。选择奇异值为全1的原因是,考察不同类型的水印图像SVD后,在正交阵 U_w 、 V_w 中包含的水印信息的容量。其中,有意义水印图像的检测结果见图4,随机水印图像的检测结果见图5。

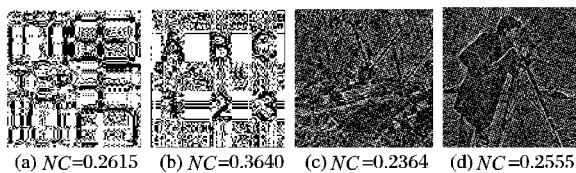


图4 置奇异值全为1时有意义水印图像的检测结果

从图4可以看出,将奇异值取为全1,也能由 U_w 和 V_w 反解出原水印图像的轮廓。 NC 值也远远大于0,约为0.25左右,这个值表征了有意义水印图像经过奇异值分解后,正交阵 U_w 、 V_w 中包含的水印信息量大小。

从图5中给出的检测结果来看,将奇异值取为全1,随机水印图像的 NC 值约为0.83~0.86,要远远高于图4中示例的 NC 值,这说明如果水印图像接近于随机分布时,奇异值的影响将变小。或者说,若嵌入的水印信息为无意义的随机信号,则SVD后水印信息将主要包含于 U_w 和 V_w 中。

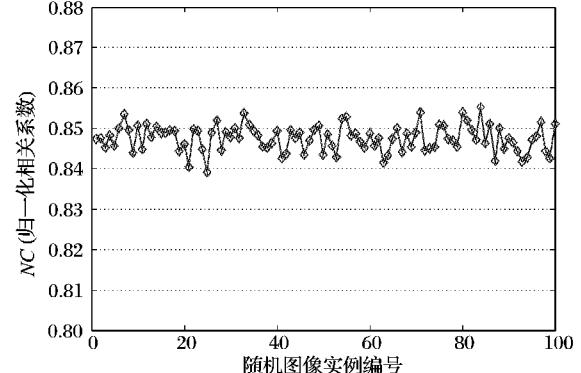


图5 置奇异值全为1时随机水印图像的检测结果

2)令 S^* 对角数据元素服从负指数分布,考察 \hat{w} 的取值情况,以及检验 w 和 \hat{w} 的相关性。选择负指数分布的原因是,对于有意义的水印图像标识,其奇异值分布接近于负指数分布,见图6。图6表明,除第一个奇异值外,其他奇异值属于缓变递减,因此取负指数分布参数 $\lambda = -0.1$ 。将奇异值替换后,有意义水印图像的检测结果见图7,随机水印图像的检测结果见图8。

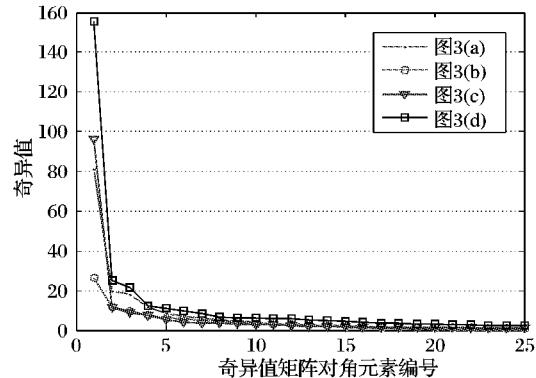


图6 有意义水印图像的奇异值分布

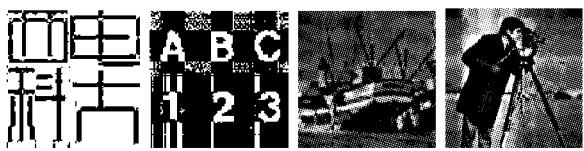


图7 置奇异值为负指数分布时有意义水印图像的检测结果

从图7的结果可以看出,将有意义水印图像的奇异值替换为服从于相同分布类型(负指数分布)的其他取值时,都能反解出高相关的水印图像以及得到高相关的 NC 值。这就是说,在实际应用场景中,可以利用 U_w 和 V_w 从其他任何有意义图像中反解出高相关的水印图像。本文第2章给出的漏洞描述实例就属于此种情形。显然,由于任何使用该类算法的用户都可以利用同一密钥在不同的图像作品中提取出他自己的水印信息,因此势必会引起作品版权隶属关系的混乱。

图8的检测结果表明,随机水印图像的奇异值替换为负指数分布值时,得到和情形1相类似的结果,这更进一步说明了水印信息为无意义的随机信号时,SVD过程将水印信息主要转移到 U_w 和 V_w 中,因此奇异值的变化对提取结果影响

较小。

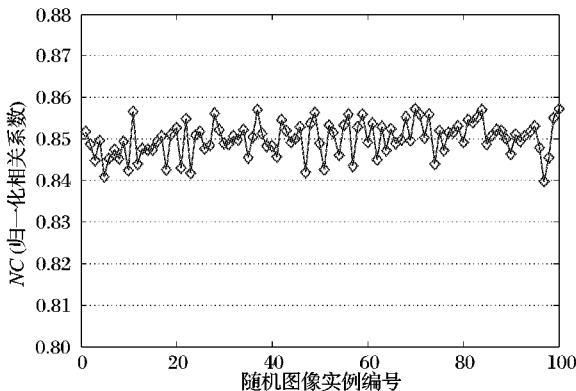


图8 置奇异值为负指数分布时随机水印图像的检测结果

3)令 S^* 对角数据元素服从负斜率分布,考察 \hat{w} 的取值情况,以及检验 w 和 \hat{w} 的相关性。选择负斜率分布的原因是,对于无意义的随机水印图像,其奇异值分布接近于负斜率分布,见图9。此处取奇异值服从 $k = -1$ 的负斜率分布。其中,有意义水印图像的检测结果见图10,无意义随机图像的检测结果见图11。

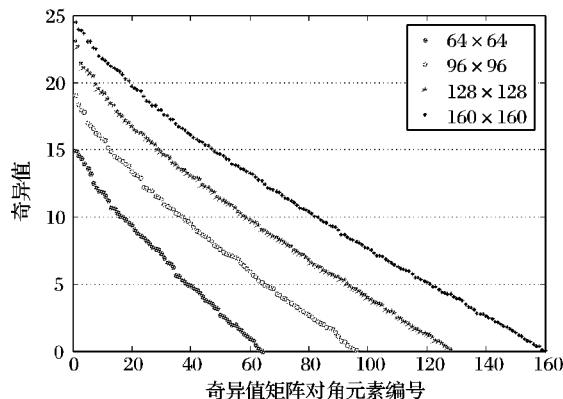
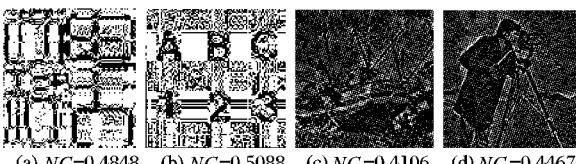


图9 服从高斯分布的随机水印图像的奇异值分布



(a) $NC=0.4848$ (b) $NC=0.5088$ (c) $NC=0.4106$ (d) $NC=0.4467$

图10 置奇异值为负斜率分布时有意义水印图像的检测结果

图10的检测结果说明,当水印图像的奇异值替换为不与原奇异值分布同类型的新值时,和情形2)相比,反解出的有意义水印图像质量变差,但要比情形1)要好。比照三种情形中奇异值分布之间的取值距离差异,可知奇异值分布类型差别越大,则检测结果越差。

图11的检测结果则映证了前述的分析。即,把随机水印图像的奇异值替换为服从相同分布类型(负斜率分布)的其他取值时,这与将有意义水印图像的奇异值替换为服从于相同分布类型(负指数分布)其他取值一样,都可以得到很好的检测值。而且由于SVD过程将随机水印信息主要转移到 U_w 和 V_w 中,因此综合起来,图11的检测结果比图5、图8都要好。

4 结语

从本文给出的详细分析结果来看,文献中给出的基于奇异值分解的图像水印算法的确存在致命漏洞,即利用在水印嵌入过程中生成的密钥可以从其他任意的图像中检测或提取

到高相关的水印信息。针对这一致命漏洞,本文给出了水印算法的伪验证定义和判定条件,然后对基于奇异值分解的水印算法进行了分析,指出在保持水印信息经奇异值分解后得到的正交阵 U_w, V_w 不变时,提取到的水印信息仅与奇异值的取值分布有关。只要代换前后的奇异值近似服从于相同的分布类型,即使具体取值不同,都能得到高相关的水印提取值。基于这一重要的分析结论,后续的水印算法设计者应不会再沿用这类基于奇异值扰动的水印算法设计思想,而需要另辟蹊径,设计其他类型的水印算法。比如文献[11]提出的通过调整正交阵 U 相邻两列系数的正负关系来嵌入水印的算法就不存在此类问题,不过该水印算法抗几何失真攻击的鲁棒性较差。因此,仍需进一步研究其他可行的水印算法。

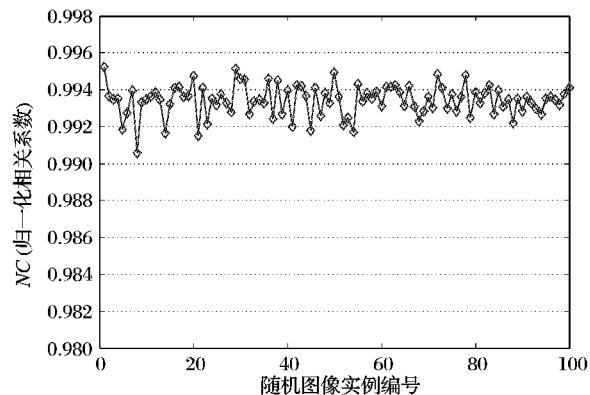


图11 置奇异值为负斜率分布时随机水印图像的检测结果

参考文献:

- [1] 刘瑞桢, 谭铁牛. 基于奇异值分解的数字图像水印方法[J]. 电子学报, 2001, 29(2): 168–171.
- [2] LIU RUI-ZHEN, TAN TIE-NIU. An SVD-based watermarking scheme for protecting rightful ownership[J]. IEEE Transactions on Multimedia, 2002, 4(1): 121–128.
- [3] CHANDRA D V S. Digital image watermarking using singular value decomposition [C]// MWSCAS02: Proceedings of 45th Midwest Symposium on Circuits and Systems. Tulsa: IEEE, 2002: 264–267.
- [4] 周波, 陈健. 基于奇异值分解的、抗几何失真的数字水印算法[J]. 中国图象图形学报, 2004, 9(4): 506–512.
- [5] 同鸣, 冯玮, 姬红兵. 一种强稳健性的抗几何攻击图像水印算法[J]. 系统仿真学报, 2008, 20(24): 6613–6616.
- [6] LIU QUAN, AI QINGSONG. A combination of DCT-based and SVD-based watermarking scheme[C]// ICSP04: Proceedings of 7th International Conference on Signal Processing. Istanbul: IEEE, 2004, 8: 873–876.
- [7] SVERDLOV A, DEXTER S, ESKICIÖGLU A M. Robust DCT-SVD domain image watermarking for copyright protection: Embedding data in all frequency [C]// EUSIPCO'05: 13th European Signal Processing Conference. Antalya: IEEE, 2005: 425–431.
- [8] GANIC E, ESKICIÖGLU A M. Robust embedding of visual watermarks using DWT-SVD [J]. IEEE Transactions on Image Processing, 2004, 4(8): 1141–1146.
- [9] PETITCOLAS F A P. Watermarking schemes evaluation[J]. IEEE Transactions on Signal Processing, 2000, 17(5): 58–64.
- [10] StirMark[EB/OL]. [2009-08-10]. http://www.petitcolas.net/fabien/software/StirMarkBenchmark_4_0_129.zip.
- [11] 张建伟, 鲍政, 王顺凤. 图像小波域分块奇异值分解的自适应水印方案[J]. 中国图象图形学报, 2007, 12(5): 811–818.