

一种新颖的用于图像内容认证、定位和恢复的半脆弱数字水印算法研究

段贵多¹, 赵 希², 李建平¹, 廖建明¹

(1. 电子科技大学计算机科学与工程学院, 四川成都 610054;

2. Department of Computing, University of Surrey, Guildford, GU2 7XH, UK)

摘 要: 本文提出了一种半脆弱, 分块和基于内容的数字水印技术. 该算法可以准确的实现篡改区域的认证, 定位和恢复. 算法基于独立分块技术将用于认证的水印比特嵌入到每个块的 Slant 变换域的中频区域. 嵌入过程是基于我们通过实验发现大部分的 Slant 中频系数的正负符号在非恶意操作前后保持不变这一性质. 在恢复系统中, 恢复比特来源于原图压缩后的数据并将此数据嵌入到图像的最低有效位以实现自恢复. 认证度由虚警检测率和误警检查率测定. 仿真实验表明我们的算法能够准确的检测和定位出篡改区域并能实现篡改区域的近似恢复. 另外, 与基于 DCT 和 PST 变换的算法相比, 我们的算法能够更有效的抵抗一些恶意和非恶意操作同时实施的操作.

关键词: 半脆弱水印; 图像认证; 恢复

中图分类号: TP309.2 **文献标识码:** A **文章编号:** 0372-2112 (2010) 04-0842-06

A Novel Semi-fragile Digital Watermarking Algorithm for Image Content Authentication, Localization and Recovery

DUAN Gui-duo¹, ZHAO Xi², LI Jian-ping¹, LIAO Jian-ming¹

(1. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Sichuan Chengdu, 610054, China;

2. Department of Computing, University of Surrey, Guildford, GU2 7XH, UK)

Abstract: A semi-fragile, block-wise and content-based watermarking method for tamper detection and recovery is presented in this paper. The non-overlapping blocks are used and the watermark bits for authentication are embedded into the middle frequency region of each block in the ST Slant Transform domain. The embedding process is based on the discovery which the sign of the most ST coefficients maintain invariant. For the recovery mechanism, the recovery bits generated from the compressed original image are embedded into the least significant bits (LSB) of the watermarked image. The degree of the authenticity is measured by the false positive detection rate and false negative detection rate. Simulation results demonstrated that our method is able to accurately detect and localize the tampered region as well as approximately recover it. Furthermore, as compare with the DCT and PST based schemes, our proposed method obtains better performance when both malicious and non-malicious manipulations are applied together.

Key words: semi-fragile watermark; image authentication; recovery

1 引言

过去的十多年里, 由于 Internet 和软件工具的迅速发展使得网络信息的复制和修改变得极其容易. 如何保护这些信息内容的完整性和真实性成为了当前迫切需要解决的问题之一. 传统的数字签名技术虽然能够达到内容认证的目的, 但数字签名作为附加信息随原作品信息传递的方式, 使得作品信息一旦发生格式改变, 签名就很容易丢失, 从而造成认证的失败. 更重要的是数字

签名无法实现篡改区域的定位和恢复, 而知道篡改位置和内容具有实际应用的价值.

脆弱和半脆弱水印技术是数字签名技术的一个有效补充, 他们在多媒体信息内容认证, 定位和恢复中已发挥了重要的作用. 脆弱水印技术^[1-3]是一种最敏感的水印技术, 它不允许作品信息有任何的改动, 甚至是一个比特的改动. 然而, 随着由传输和存储引起的轻微的信号处理操作, 诸如 JPEG 压缩, 加噪, 被认为是可接受而且需要的操作后, 半脆弱水印更适合于实际应用的

假设任意 8×8 的原图和被压缩图的任意一个 Slant 系数分别记为: x 和 x' ,

如果操作前 $x \geq 0$ 和操作后 $x' \geq 0$ 成立,

或者操作前 $x < 0$ 和操作后 $x' < 0$ 成立,

则我们认为这个系数的正负符号在 JPEG 压缩操作前后是不变的。

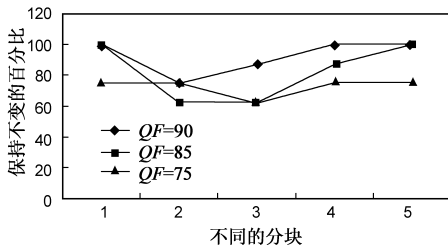


图3 'San diego'中系数保持正负不变的个数的百分比

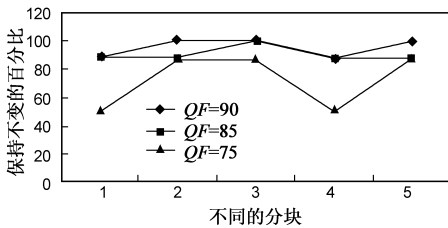


图4 'Bridge'中JPEG系数保持正负不变的个数的百分比

上述实验在 6 幅 512×512 标准测试图 (Lena, Baboon, Ship, Trucks, Bridge 和 San Diego) 中进行, 图 3 和图 4 给出的是 'San Diego' 和 'Bridge' 的情况. 图 3 中的纵坐标中百分比 P 定义如下:

$$P = \% \frac{\text{某个分块中系数正负保持不变的数目}}{\text{分块中所有的系数个数}}$$

横坐标中 5 个数字分别代表 5 个随机选择的分块. 图 4 中情况类似. 关于 Gussian 加噪后的表现情况见图 5 和图 6, 其计算方式与上类似.

从图 3 和图 4 可以看出, 随着 QF 的降低, 压缩比增大, 系数符号正负保持不变的百分比也随之下降. 例如, QF = 75 的时候, 百分比也至少保持在 50% 以上. 图 5 和图 6 展示的是 Gussian 加噪后的情况, 随着噪音强度的增强, 结果比较接近, 说明这几个强度参数效果相近. 虽然图中出现了一些波动, 但这是随机选择的结果. 总之, 大部分的系数的正负符号在操作前后保持不

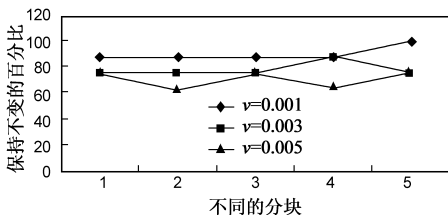


图5 'San diego'中Gussian加噪后系数保持正负不变的个数的百分比

变, 利用这个性质进行系数调节期望得到更好的鲁棒性.

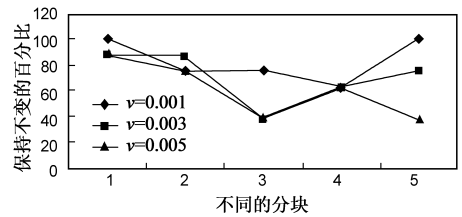


图6 'Bridge'中Gussian加噪后系数保持不变的个数的百分比

2.2.2 认证和恢复水印的嵌入过程

本算法中水印为一伪随机二值序列 $\{-1, 1\}$, 图像首先被分为 8×8 大小的互不重叠的分块, 然后将每个分块的每个像素值的最低有效位置零, 接着对每个小块实施 Slant 变换. 水印通过调节中频的 Slant 系数 x 与事先设置的阈值 τ 之间的大小关系完成嵌入, 其嵌入细节如下:

当水印比特为 1 的时候, 如果 $x \geq \tau$, 则不做任何操作, 否则, 将 x 替换成阈值 α ;

当水印比特为 -1 的时候, 如果 $x < -\tau$, 则不做任何操作, 否则, 将 x 替换成阈值 $-\alpha$;

其中, x 为原始图像的 Slant 变换系数, x' 为嵌入水印后的系数, w 为水印比特, $\tau > 0$ 为控制嵌入水印后图像质量的阈值, $\alpha \in [\frac{\tau}{2}, \tau]$ 为一个常量. 最后, 利用重构算法得到含水图片并丢弃最低有效位.

对于恢复机制, 我们采用的是类似于文献[4]的恢复机制, 不过将 DCT 变换替换为 Slant 变换. 恢复的基本思想是将原图的一个压缩版本嵌入到图像中. 简单地说, 算法将每个的独立分块进行 Slant 变换, 并量化变换后的系数. 接着, 将利用 Zigzag 扫描后获得前 11 个 Slant 变换系数进行压缩. 压缩过程所采用的是文献[4]所用的 JPEG 量化表(见图 7), 这主要是考虑到 Slant 变换系数的动力范围比 DCT 小些. 将量化后的系数经由图 8 编码为 64bit 并作为恢复信息随机嵌入到 LSB 中. 为了

1.6	1.1	1.0	1.6	2.4	0	0	0
1.2	1.2	1.4	0	0	0	0	0
1.4	1.3	0	0	0	0	0	0
1.4	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

图7 前11个Slit系数的量化表

增加自恢复算法的安全性,每个块中被编码的比特都被加密.进一步,为提高恢复算法的鲁棒性,算法对每个分块实施置乱变换,即某个块对应的恢复信息嵌入到另一个块中,而这层映射关系由密钥控制.但需注意的是由于篡改区域周围的分块也容易被篡改,所以某块与其映射块之间应有一定的距离.

7	7	7	5	4	0	0	0
7	6	5	0	0	0	0	0
6	5	0	0	0	0	0	0
5	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

图8 被用于为前11个SlT系数编码的比特

水印算法的流程图可见图 9,其由两部分组成.一部分为原图的前 7 个位平面携带这认证水印,另一部分为最低有效位携带着为恢复而用的原图的压缩版信息.

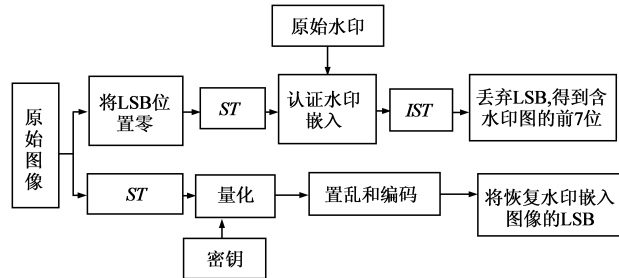


图9 认证和恢复水印的嵌入过程

3 水印提取,认证和恢复算法

根据式(2),提取算法比较简单.同样的,先将含水印的图像分为大小为 8×8 的互不重叠的块,并将每个分块的像素值的 LSB 清零,随后对每个分块实行 Slant 变换,针对每块的中频部分的系数按照式(2)即可提取出水印:

$$w' = \begin{cases} 1 & (y \geq 0) \\ -1 & (y < 0) \end{cases} \quad (2)$$

其中 y 为检测图像的中频部分系数, w' 为提取的水印.在认证过程中,将提取的水印和原始水印进行相关性检测,其计算方式如式(3):

$$\rho(w, w') = \frac{\sum w(n)w'(n)}{\sqrt{\sum w^2(n)\sum w'^2(n)}} \quad (3)$$

某个块的认证基于以下判别标准:如果 $\rho \geq \lambda$,则认为

该块通过认证,否则认为该块被篡改,被篡改的块用黑色表示(见图 11(c)、(g)).确定某个块被篡改后,则实施恢复算法.首先利用映射变换确定篡改区域的恢复信息所在的块,从中提取 64 位 LSB 并利用密码进行解码和反量化变换获得嵌入的恢复比特,以此实现篡改区域的恢复.有关提取,认证和恢复算法的流程见图 10.

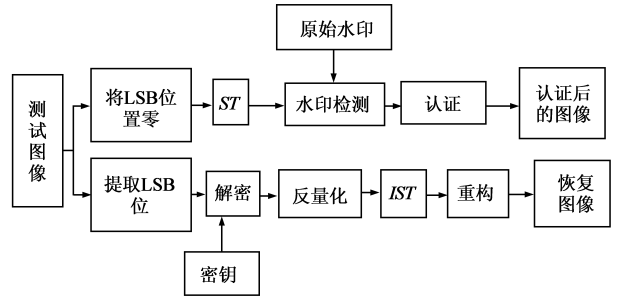


图10 水印提取,认证和图像恢复

4 仿真实验

本节利用一系列标准的测试图片(Lena, Baboon, Ship, Trucks, Bridge, San Diego 和 Singapore),大小为 512×512 ,通过一系列的对比实验评估了本算法的性能.

与本算法做比较的算法为基于 DCT^[4]和基于 PST^[7]的两种算法.在所有算法中,水印都被嵌入各自对应的频域中的 8 个中频系数,参见图 2.为了公平比较,每种水印算法中的嵌入强度都调节到保证嵌入后水印图像的 PSNR 值大约在 33db 左右.算法的性能通过虚警检测率 (P_{FP}),误警检测率 (P_{FN}),和平均检测率 (P_D)测定,其定义为:

$$P_{FP} = \% \text{ 未被篡改的分块被误检为被篡改}$$

$$P_{FN} = \% \text{ 被篡改的分块被误检为未被篡改}$$

$$P_D = \% (100 - \frac{P_{FP} + P_{FN}}{2})$$

另外,在所有实验中,检测阈值 λ 被设置为 0.5,一个实验测试值. λ 的选择还考虑到前面图 3 到图 6 的测试中,在各个强度的非恶意操作下,50% 的系数的正负符号能够保持不变.

4.1 单独抵抗复制粘贴操作的性能表现

在第一类实验中,我们将测试水印算法单独抵抗复制粘贴操作的性能表现.在这类攻击测试中,复制粘贴操作用的是从相同图像中随机选择的部分图像.表 1 展示的是采用不同的水印密码重复 10 次复制和粘贴操作(20% 的篡改)的实验结果的平均值.从表 1 可以得出,三种水印算法的表现相近,不过基于 Slant 变换的算法效果更佳:在遭受 20% 的篡改攻击下,无虚警率,误检率在 10% 以下.图 11 展示了一些被复制粘贴攻击后的实例.图 11(a)、(e)为含水印图,(b)中有 3 个区域被

篡改, (f) 中有一个区域被篡改, (c)、(g) 为实现认证后的图, (d)、(h) 为恢复后的图. 表 1 和图 11 表明我们的算法可以对篡改部位实现准确的认证, 定位和较高质量恢复.

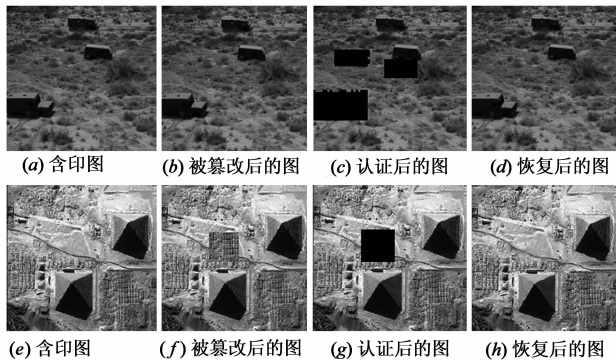


图 11 实例

表 1 复制粘贴攻击表现(20%篡改)

测试 图像	ST			DCT			PST		
	P_{FP}	P_{FN}	P_D	P_{FP}	P_{FN}	P_D	P_{FP}	P_{FN}	P_D
Lena	0	9.06	95.47	0.05	9.68	95.13	0	9.95	95.02
Baboon	0	9.41	95.29	0.19	9.47	95.17	0	10.11	94.95
Ship	0	9.99	95.01	0.08	10.26	94.83	0	9.78	95.11
Trucks	0	9.47	95.26	0.01	9.56	95.21	0	9.47	95.26
Bridge	0	9.38	95.31	0.17	10.01	94.91	0	9.57	95.21
San Diego	0	9.65	95.18	0.04	9.69	95.13	0	9.06	95.47
Avg.	0	9.49	95.25	0.09	9.78	95.06	0	9.66	95.17

4.2 同时抵抗复制粘贴连同 JPEG 压缩, 高斯加噪操作的性能表现

由于在实际应用中, 图像更可能的是同时经历恶意和非恶意的操作, 表 2 到表 5 展示了三种算法在面对同时经历复制粘贴这一恶意操作和 JPEG 压缩, Gaussian 加噪这两种非恶意操作时各自不同的性能表现. 当压缩品质为 $QF = 85$, 我们的算法比其他两种算法效果稍稍好些. 但是随着压缩比例的增加, 三种算法之间的表现就有所不同, 基于 DCT 和 PST 的算法在抵抗 JPEG 压缩攻击时比本算法效果好些. 对于加噪的情况, 从这些表里面我们还发现, 对于不同的图像得到了不同的结果, 纹理丰富的图像的结果要相对好些, 诸如 Baboon 和 San_Diego.

总之, 基于 Slant 变换的水印算法在抗击复制粘贴这一恶意攻击的时候比其他两种算法准确性高, 而在抗击加噪攻击时候鲁棒性更强, 不过对于 JPEG 压缩攻击稍显逊色, 这主要是因为 JPEG 压缩是基于 DCT 的算法.

表 2 复制粘贴(20%篡改攻击)连同 JPEG 压缩攻击($QF = 85$)表现

Test image	ST			DCT			PST		
	P_{FP}	P_{FN}	P_D	P_{FP}	P_{FN}	P_D	P_{FP}	P_{FN}	P_D
Lena	0	8.58	95.71	0.13	9.85	95.01	0	9.99	95.01
Baboon	0.01	9.63	95.18	0.35	9.44	95.1	0	9.9	95.05
Ship	0	9.47	95.26	0.13	10.05	94.91	0	9.35	95.32
Trucks	0	9.89	95.05	0.07	9.58	95.17	0	9.39	95.31
Bridge	0.05	9.56	95.19	0.4	10.13	94.73	0.03	9.56	95.21
San Diego	0	9.55	95.22	0.15	9.8	95.02	0	9.57	95.21
Avg.	0.01	9.45	95.27	0.21	9.81	94.99	0.01	9.63	95.19

表 3 复制粘贴(20%篡改攻击)连同 JPEG 压缩攻击($QF = 75$)表现

Test image	ST			DCT			PST		
	P_{FP}	P_{FN}	P_D	P_{FP}	P_{FN}	P_D	P_{FP}	P_{FN}	P_D
Lena	21.07	9.58	84.67	0.75	9.07	95.09	0	9.49	95.26
Baboon	9.69	9.84	90.23	1.58	9.13	94.64	0.01	10.04	94.98
Ship	20.21	9.32	85.24	0.96	9.93	94.56	0.01	8.86	95.56
Trucks	15.04	9.22	87.87	0.88	9.55	94.78	0	9.78	95.11
Bridge	12.54	9.80	88.83	1.73	10.29	93.99	0.32	9.78	94.95
San Diego	8.94	9.99	90.53	1.40	9.43	94.59	0	9.63	95.18
Avg.	14.58	9.63	87.90	1.22	9.57	94.61	0.06	9.60	95.17

表 4 复制粘贴(20%篡改攻击)连同 Gaussian 加噪攻击($v = .003$)表现

Test image	ST			DCT			PST		
	P_{FP}	P_{FN}	P_D	P_{FP}	P_{FN}	P_D	P_{FP}	P_{FN}	P_D
Lena	10.21	9.48	90.15	15.60	9.65	87.38	11.72	9.20	89.54
Baboon	10.39	9.40	90.11	16.14	8.59	87.63	11.47	10.50	89.01
Ship	10.58	9.56	89.93	16.23	10.58	86.59	11.77	9.36	89.44
Trucks	10.53	8.79	90.34	15.84	9.81	87.18	12.17	9.44	89.20
Bridge	10.19	10.46	89.67	16.18	9.89	86.96	11.60	10.26	89.07
San Diego	10.26	9.93	89.90	15.50	10.66	86.92	11.45	9.97	89.29
Avg.	10.36	9.60	90.02	15.92	9.86	87.11	11.70	9.79	89.26

表 5 复制粘贴(20%篡改攻击)连同 Gaussian 加噪攻击($v = .005$)表现

Test image	ST			DCT			PST		
	P_{FP}	P_{FN}	P_D	P_{FP}	P_{FN}	P_D	P_{FP}	P_{FN}	P_D
Lena	13.52	8.83	88.82	21.97	10.62	83.70	16.94	9.61	86.73
Baboon	13.49	9.52	88.49	21.34	9.93	84.36	15.80	10.46	86.87
Ship	14.21	9.81	87.99	22.34	10.62	83.52	16.96	9.85	86.60
Trucks	13.46	10.18	88.18	21.76	9.85	84.20	16.87	9.52	86.81
Bridge	13.59	10.22	88.10	21.89	9.52	84.29	16.90	9.08	89.07
San Diego	13.50	10.26	88.12	21.66	10.91	83.72	16.64	9.56	86.90
Avg.	13.63	9.80	88.28	21.83	10.24	83.97	16.69	9.68	86.82

5 结论

本文提出了一种基于 Slant 变换的半脆弱水印算法, 该算法可以用于图像内容的认证, 定位和恢复. 本算法将认证水印嵌入 Slant 变换后的中频系数中, 并利用 LSB 实现篡改区域的自恢复. 大量的对比实验表明 Slant 变换在用于提高篡改检测的精确性的有效性和潜力. 当单独复制粘贴攻击时, 当篡改率达到 20% 时候, 本算法在无虚警检测率的情况下, 误警检测率也是最低的(10% 以下), 这是其他两种算法不可及的. 另外, 该算法在抵抗复制粘贴联合 Gaussian 加噪攻击方面表现

更优秀,不过在抵抗复制粘贴联合较强的 JPEG 压缩攻击方面有所不如。

但是,基于 LSB 的恢复算法无法较好的抵抗 JPEG 压缩,因此下一步的工作是我们将利用一种最新的图像多尺度变换——非冗余 Contourlet 变换^[17]后的系数关系^[18]实现鲁棒的恢复算法。

致谢:感谢英国 University of Surrey, Department of Computing 的 Ho 教授和 Pankajakshan 博士为本论文所做的工作。

参考文献:

- [1] 张宪海,杨永田.基于脆弱水印的图像认证算法研究[J].电子学报,2007,35(2):34-39.
Zhang Xian-hai, Yang Yong-tian. Image authentication scheme research based on fragile watermarking[J]. Acta Electronica Sinica, 2007, 35(2): 34-39. (in Chinese)
- [2] Lin P L, Hsieh C K, Huang P W. A hierarchical digital watermarking method for image tamper detection and recovery[J]. Pattern Recogn. Lett., 2005, 38(12): 2519-2529.
- [3] Chang C C, Hu Y S, Liu T C. A Watermarking-based image ownership and tampering authentication scheme[J]. Pattern Recogn. Lett., 2006, 27: 439-446.
- [4] Fridrich J, Goljan M. Images with self-correcting capabilities [A]. IEEE Int. Conf. on Image Processing, vol. 3[C]. Kobe, Japan, Oct, 1999. 792-796.
- [5] Lin CY, Chang S F. Semi-fragile watermarking for authenticating JPEG visual content [A]. Proc. SPIE, Security and Watermarking of Multimedia Content II [C], San Jose, CA, Jan. 2000: 140-151.
- [6] Al-Mualla M E. Content-Adaptive Semi-Fragile Watermarking for Image Authentication [A]. 14th IEEE Int. Conf. on Electronics, Circuits and Systems [C]. Marrakech, Morocco Dec. 2007. 1256-1259.
- [7] Ho A T S, Zhu X, Guan Y. Image content authentication using pinned sine transform [J]. EURASIP Journal on Applied Signal Processing, 2004, 14: 2174-2184.
- [8] 胡玉平,陈志刚.用于图像认证的小波域半易损水印算法[J].电子学报,2006,34(4):653-657.
Hu Yu-ping, Chen Zhi-gang. Wavelet domain semi-fragile watermarking algorithm for image authentication [J]. Acta Electronica Sinica, 2006, 34(4): 653-657. (in Chinese)
- [9] Matro K, Sun Q, Chang S, et al. New semi-fragile image authentication watermarking techniques using random bias and nonuniform quantization [J]. IEEE Trans. Multimedia, 2006, 8(1): 32-45.
- [10] Tsai M J, Chien C C. Authentication and recovery for wavelet-based semi-fragile watermarking [J]. Optical Engineering, 2008, 47(6): 1-10.
- [11] Pratt W, Chen W H, Welch L. Slant transform image coding

[J]. IEEE Trans. Commun., 1974, 22(8): 1075-1093.

- [12] Zhu X, Ho A T S. A slant transform watermarking for copyright protection of satellite images [A]. Fourth Pacific Rim Conference on Multimedia, Vol 2 [C]. Dec. 2003. 1178-1181.
- [13] 张静,张春田.用于 JPEG2000 图像认证的半脆弱数字水印算法[J].电子学报,2004,32(1):157-160.
ZHANG Jing, ZHANG Chun-tian. Semi-fragile watermarking for JPEG 2000 image authentication [J]. Acta Electronica Sinica, 2004, 32(1): 157-160. (in Chinese)
- [14] 杨义先,钮心忻.数字水印理论与技术[M].北京:高等教育出版社,2006.156-161.
Yang Y, Niu X. Theory and Applications of Digital Watermarking [M]. Beijing: Higher Education Press, 2006. 156-161. (in Chinese)
- [15] Wang M S, Chen W C. A majority-voting based watermarking scheme for color image tamper detection and recovery [J]. Pattern Recogn. Lett., 2007, 29: 561-570.
- [16] Hou Z, Xu N, Chen H, et al. Fast slant transform with sequence increment and its application in image compression [A]. Proceedings of 2004 International Conference on Machine Learning and Cybernetics, Vol. 7 [C]. Shanghai, China. Aug. 2004. 4085-4089.
- [17] Do M N, Vetterli M. The contourlet transform: an efficient directional multiresolution image representation [J]. IEEE Trans. Image Proc., 2005, 14: 2091-2106.
- [18] Duan G, Li J, Huang, T. A Robust Watermarking Algorithm Based on Significant-tree in Contourlet Domain [J]. High tech Lett, 2008, 14(1): 67-71.

作者简介:



段贵多 女,1981年出生于四川荣经,工学博士,曾经于2007年9月到2008年9月在英国萨里大学计算系做学术研究.以第一作者身份在国内外期刊和会议集发表论文7篇,主要研究方向:多尺度几何分析、数字水印、图像处理.
E-mail: duanguiduo@163.com

赵希 男,1980年出生,英籍华人,英国萨里大学计算系博士生,已在国外杂志和会议集上发表论文数篇,主要研究方向:数字水印. E-mail: x.zhao@surrey.ac.uk

李建平 男,1964年出生于湖南祁阳,电子科技大学教授、博导、国际小波分析应用研究中心主任,为国际上小波分析与信号处理领域较为活跃的专家.主要研究方向:小波分析、模式处理和图像处理.

廖建明 男,1963年出生于四川广安,电子科技大学教授,1982年7月毕业于重庆大学,曾于1989年3月和1997年10月在日本广岛大学计算机系和日本千叶大学计算机系进行访问研究.主持和参加了国家863和省级多项科研项目,并多次获奖.参与编写论文和著作多部.主要研究方向:计算机测控技术和计算机应用.