

文章编号 1004-924X(2010)11-2467-06

数字图像与数码相机噪声相关性的偏态分布

崔夏荣

(武夷学院 电子工程系, 福建 武夷山 354300)

摘要: 为了提高利用噪声相关性鉴别数字图像真伪的正确性, 探索了描述噪声相关性偏态分布的最佳模型。首先, 通过理论分析和大量实验验证了在 $(0, 1]$ 区间 beta 分布、gamma 分布和对数正态分布可用于描述噪声相关性的偏态分布。然后, 利用这 3 种分布的概率密度函数模拟实际的噪声相关性偏态分布的概率密度函数曲线, 概率密度函数曲线的特征和最小错误率的大小说明了采用对数正态分布描述噪声相关性偏态分布的效果最佳, 从而提出用对数正态分布描述噪声相关性偏态分布的模型。实验结果表明, 与采用广义 chi 平方分布的模型相比, 采用该模型可使最小错误率降低 60% 以上, 证明了采用正确的模型描述噪声相关性偏态分布是降低鉴别错误率的有效途径。

关键词: 数字图像; 数码相机; 噪声相关性; 偏态分布; 概率密度函数

中图分类号: TP391.41; TB853.1 **文献标识码:** A **doi:** 10.3788/OPE.20101811.2467

Asymmetrical distribution of noise correlation between digital image and digital camera

CUI Xia-rong

(Department of Electronic Engineering, Wuyi University, Wuyishan 354300, China)

Abstract: The best model to describe the asymmetrical distribution of noise correlation was explored to improve the accuracy of identifying digital image authenticity by noise correlation. By theoretical analysis and experiments, it was indicated that the beta distribution, gamma distribution and the logarithmic normal distribution could be used to describe the asymmetrical distribution of noise correlation within $(0, 1]$. Then, three kinds of probability density functions were used to simulate the probability density function curves of the actual asymmetrical distribution of noise correlation. On the basis of the curve modality of probability density function and the value of the least false rate, it was explained that using logarithmic normal distribution to describe the asymmetrical distribution of noise correlation could obtain the best result. Thus a model by using the logarithmic normal distribution to describe the asymmetrical distribution of noise correlation was proposed. Experiment results show that this model can reduce the least false rate over 60% as compare with the model of using generalized chi square distribution, which proves that using a correct model to describe the asymmetrical distribution of noise correlation is an effective approach to reduce the false rate.

Key words: digital image; digital camera; noise correlation; asymmetrical distribution; probability density function

收稿日期: 2010-05-18; 修订日期: 2010-08-12.

基金项目: 福建省教育厅科技基金资助项目 (No. JA08245)

1 引言

在被动式数字图像真伪鉴别中,利用数字图像与数码相机噪声相关性的鉴别方法的最大特点是鉴别灵敏度非常高,因此有很强的应用潜力和很高的研究价值。由于该方法在鉴别实践中一般采用贝叶斯的决策理论,因而必须预知被鉴别样本的概率分布类型。大量的实验结果表明,数字图像与数码相机噪声相关性的分布有对称分布和非对称(偏态)分布两种类型^[1-2]。对称分布用广义高斯分布(Generalized Gaussian Distribution—GGD)描述可以获得更逼近的误差估计^[1]。文献[3-5]没有明确指出噪声相关性存在偏态分布,对噪声相关性的分布都采用 GGD 描述,只在文献[5]中说明了主要采用 GGD 描述,这一点值得注意。本文仅讨论噪声相关性的偏态分布。

目前,虽然描述偏态分布的概率密度函数(Probability Density Function, PDF)有很多,如 β 分布、 Γ 分布、 F 分布、Relley 分布、chi 平方分布、Rician 分布、对数正态分布等等,但是尚无文献明确指出噪声相关性的偏态分布符合哪一种分布类型。文献[1]在(0,1]区间构造了一个广义 chi 平方分布的 PDF 来描述这种偏态分布,获得了较好的效果,但是它没有说明采用该 PDF 效果是最佳的。文献[6-8]介绍了根据最大熵原理,在已知或假设随机变量满足某些约束条件的前提下,利用拉格朗日(Lagrange)乘法推导出各种 PDF 的方法。然而对于噪声相关性的偏态分布,除了已知其 PDF 对 x 的积分等于 1 之外,目前还没有任何文献介绍其满足其它的约束条件,因此应用该方法无法推导出它满足上述的哪一种分布。

由于噪声相关性偏态分布的定义域为(0,1]^[1],因此上述的 PDF 不可能都适合于描述噪声相关性的偏态分布。由于不同 PDF 的曲线不同,选用不同 PDF 时所得到的鉴别精度也就不同。为了提高利用噪声相关性鉴别数字图像真伪的正确性,探讨上述哪一个偏态分布的 PDF 最适合用于描述噪声相关性的偏态分布是很有意义的。

本文在文献[1-3]的研究基础之上,按照理论与实验相结合的原则,应用 PDF 曲线的模拟和最小错误率估计的方法,讨论了最适合描述噪声相关性偏态分布的 PDF 问题,与文献[1]相比较,希

望获得更低的鉴别错误率。

2 噪声相关性的检测原理

文献[9]详细介绍了数码相机内部噪声的来源及噪声模型。本文提取数字图像的残留噪声(或称样本噪声)与计算噪声相关性的方法均与文献[1]相同,图像被检测区域的尺寸均取 1 024 pixel×1 024 pixel,且均为图像的中心区域。

利用小波滤波器提取图像 I 的样本噪声 n 的过程可表示为

$$n = I - F(I), \quad (1)$$

式中 $F(I)$ 表示小波滤波器的输出。相机的参考模式噪声 n_c 通过平均多幅图像的样本噪声 n 的方法得到

$$n_c = \frac{1}{N} \sum_{k=1}^N [I_k - F(I_k)], k = 1, 2, \dots, N, \quad (2)$$

数字图像与数码相机的噪声相关系数 r 的大小由下式得到

$$r = \frac{E((n - E(n)) \times (n_c - E(n_c)))}{\sqrt{E[(n - E(n))^2]} \sqrt{E[(n_c - E(n_c))^2]}}, \quad (3)$$

式(3)中 $E(\cdot)$ 表示均值, $E(n)$ 和 $E(n_c)$ 分别是 n 和 n_c 的像素亮度均值。

3 噪声相关性的偏态分布

图 1 是松下 DMC-FX7 相机的 355 幅分辨率为 2 560 pixel×1 920 pixel 的原始图像与相机的噪声相关性的分布图,显然其为偏态分布。在 7 000 幅图像的数据库中, Olympus、Canon 等品牌相机的噪声相关性也出现了这种偏态分布,其

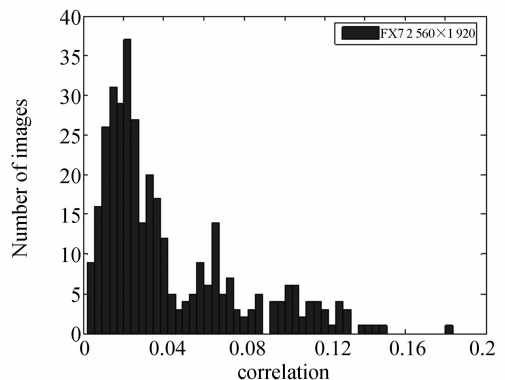


图 1 噪声相关性的偏态分布

Fig. 1 Asymmetrical distribution of noise correlation

所占的相机比例超过了 40%。文献[1]分析了影响噪声相关性分布特性的各种因素,并指出只有 j 相机拍摄的图像与 j 相机的噪声相关性 $r_{j,j}$ 才会出现偏态分布,当图像被检测区域的尺寸为 $1\ 024\ \text{pixel} \times 1\ 024\ \text{pixel}$ 时, $r_{j,j}$ 的定义域为 $(0, 1]$ 。

4 噪声相关性偏态分布的 PDF

把 $r_{j,j}$ 看成一个连续型变量 x ,由文献[10-11]可知,由于 chi 平方分布的均值 μ 和方差 δ^2 的关系为

$$\delta^2 = 2\mu, \quad (4)$$

Relley 分布的均值 μ 和方差 δ^2 的关系为

$$\frac{\mu^2}{\delta^2} = \frac{\pi}{4-\pi}, \quad (5)$$

即 chi 平方分布、Relley 分布的均值 μ 和方差 δ^2 满足某种特定的常数关系。实验结果表明,噪声相关性的偏态分布一般不满足这种关系,因此这 2 个 PDF 不能用于描述噪声相关性的偏态分布。

同理,通过理论分析和实验可知,在 $(0, 1]$ 区间可以用于描述噪声相关性偏态分布的概率密度函数有 beta 分布、gamma 分布和对数正态分布。

4.1 β 分布

β (beta)分布的 x 的定义域为 $(0, 1)$,由于其定义域不包含 1,因此从严格的数学角度看它不能用于描述噪声相关性的偏态分布;但大量的实验表明 $r_{j,j}$ 一般不会达到 1,因此一般情况下可以用 β 分布描述噪声相关性的偏态分布。其 PDF 表达式为

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}. \quad (6)$$

其中参数 $\alpha > 0, \beta > 0$ 。其均值 μ 和方差 δ^2 分别为

$$\mu = \frac{\alpha}{\alpha+\beta}, \quad (7)$$

$$\delta^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}, \quad (8)$$

由式(7)和(8)可解得参数 α 和 β :

$$\alpha = \frac{\mu(\mu - \mu^2 - \delta^2)}{\delta^2}, \quad (9)$$

$$\beta = \frac{(1-\mu)(\mu - \mu^2 - \delta^2)}{\delta^2}. \quad (10)$$

4.2 Γ 分布

在自然界中服从 Γ (gamma)分布的现象不少,我国的水文界广泛用 Γ 分布来模拟水文数据, Γ 分布兼有指数分布和幂分布的特点,其 PDF 表达式为

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp(-\frac{x}{\beta}) & x > 0 \\ 0 & \text{otherwise} \end{cases}, \quad (11)$$

其中参数 $\alpha > 0, \beta > 0$, α 是形状参数, β 是尺度参数。其均值 μ 和方差 δ^2 分别为 $\alpha\beta$ 和 $\alpha\beta^2$ 。显然, $\alpha = \mu^2/\delta^2, \beta = \delta^2/\mu$ 。

比较 chi 平方分布与 Γ 分布的 PDF 可知,chi 平方分布只是 Γ 分布在 $\beta = 2, \alpha = m/2$ 时的一个特例,因此文献[1]构造的广义 chi 平方分布也是 Γ 分布的一个特例。由于广义 chi 平方分布的适用条件是 $\mu(kx) = m$ 和 $\sigma^2(kx) = 2m$,且 m 为自然数,因此 Γ 分布比广义 chi 平方分布有更广泛的适应性,选用 Γ 分布描述噪声相关性的偏态分布比选用广义 chi 平方分布更优越。

4.3 对数正态分布

对数正态分布(logarithmic normal distribution)是 x 的对数 $\ln x$ 为正态分布的概率分布,其 PDF 表达式为

$$f(x) = \begin{cases} \frac{1}{x\theta\sqrt{2\pi}} \exp[-\frac{(\ln x - \lambda)^2}{2\theta^2}] & x > 0 \\ 0 & \text{otherwise} \end{cases}, \quad (12)$$

其中参数 λ 是 $\ln x$ 的均值, θ 是 $\ln x$ 的标准差。 x 的均值 μ 和方差 δ^2 分别为

$$\mu = \exp(\lambda + \frac{\theta^2}{2}), \quad (13)$$

$$\delta^2 = [\exp(\theta^2) - 1] \exp(2\lambda + \theta^2). \quad (14)$$

由式(13)和(14)可解得参数 λ 和 θ :

$$\lambda = \ln \mu - \frac{1}{2} \ln(\frac{\delta^2}{\mu^2} + 1), \quad (15)$$

$$\theta^2 = \ln(\frac{\delta^2}{\mu^2} + 1). \quad (16)$$

4.4 概率密度函数曲线的模拟

在实验数据的统计中,通常得到的是 x 的平均值 μ 和方差 δ^2 。图 1 中 $r_{j,j}$ 的均值 $\mu = 0.042\ 4$, 方差 $\delta^2 = 0.001\ 2$,利用上述 3 个 PDF 画出相应的 PDF 曲线,如图 2 所示。

在图 2 中,实线型曲线是对数正态分布的

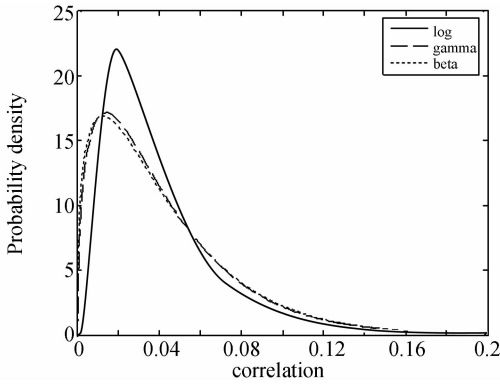


图 2 DMC-FX7 相机的 3 种 PDF 的曲线

Fig. 2 Three kinds of PDF curves for camera DMC-FX7

PDF 曲线, 双划线型曲线是 Γ 分布的 PDF 曲线, 虚线型曲线是 β 分布的 PDF 曲线。可见, 3 条曲线并不完全重合, 3 条曲线的峰值位置也不同。比较图 1 和图 2 可知, 从两者的分布形态和峰值位置的角度考虑, 选用对数正态分布描述图 1 的偏态分布效果最佳。为了能更直观地比较两者的相似度, 把图 1 的噪声相关性分布和图 2 的 PDF 曲线组合在图 3 中。由于两者纵轴的物理意义不同, 为了方便观察, 在纵轴方向对 PDF 曲线进行了适当的线性处理。由图 3 可见, 对数正态分布 PDF 曲线的形态与实际的噪声相关性分布形态的相似度最高。

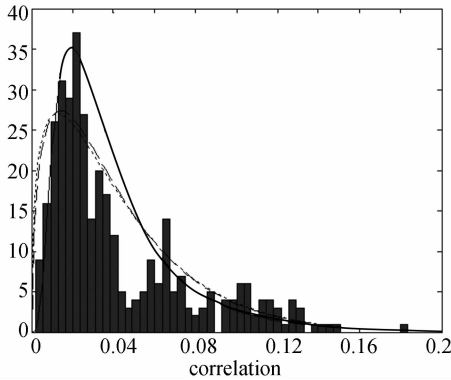


图 3 DMC-FX7 相机的相关性与 PDF 曲线

Fig. 3 Correlation and PDF curves for camera DMC-FX7

为了验证采用对数正态分布描述噪声相关性偏态分布的最佳性, 做了大量的实验。图 4 是 Olympus C5060 相机的 434 幅图像的 $r_{j,j}$ 分布分别采用对数正态分布、 Γ 分布、 β 分布描述时得到的 3 条 PDF 曲线, 其 $\mu = 0.1028, \delta^2 = 0.0028$ 。比较图 4 与文献 [1] 的图 3 可知, 对数正态分布 PDF 曲线“左尾”的形态与峰的形态都更贴近实

际的分布形态。大量的实验结果表明, 对于噪声相关性的偏态分布采用上述 3 个 PDF 描述时均是对数正态分布的效果最佳。因此本文推荐采用对数正态分布来描述噪声相关性的偏态分布。

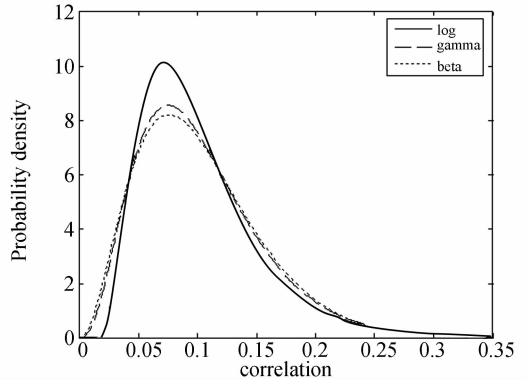


图 4 C5060 相机的 3 种 PDF 的曲线

Fig. 4 Three kinds of PDF curves for camera C5060

5 在数字图像来源鉴别中的应用

为了验证用对数正态分布描述噪声相关性偏态分布的优越性, 应用最小错误率的 Bayes 决策规则^[1], 对 Olympus C5060、C460 和 Canon 400 3 台相机拍摄的原始图像进行图像来源鉴别实验。图 5 是 C460 相机拍摄的 335 幅分辨率为 $2288 \text{ pixel} \times 1712 \text{ pixel}$ 的图像与相机的噪声相关性 $r_{j,j}$ 的分布图, 其为偏态分布, 其 $r_{j,j}$ 的均值 $\mu = 0.0362$, 方差 $\delta^2 = 3.0493 \times 10^{-4}$ 。

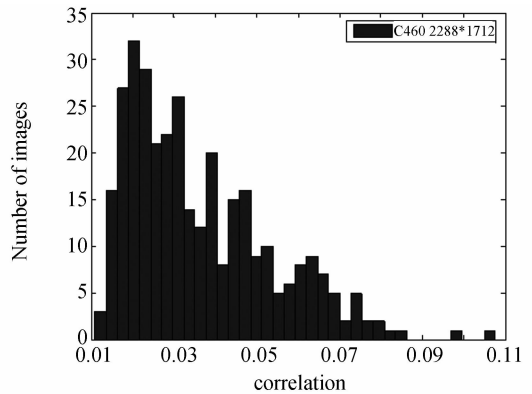


图 5 c460 的图像与 c460 的相关性

Fig. 5 Correlation of images from c460 with c460

图 6 是其它 9 种相机拍摄的 2700 幅的图像与 C460 相机的噪声相关性 $r_{j,B}$ 的分布图, 其为对称分布, 其 $r_{j,B}$ 的均值 $\mu = 0.0010$, 方差 $\delta^2 =$

4.438 4 × 10⁻⁶, 偏差 θ = 0.001 6。

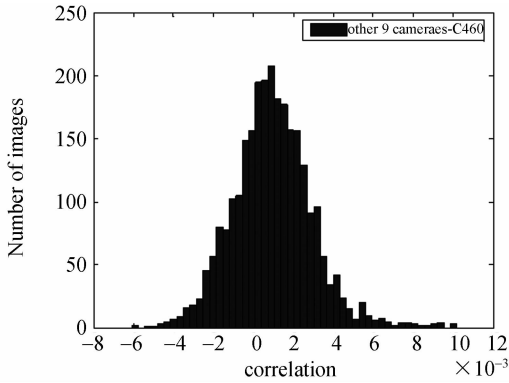


图 6 其它 9 种相机的图像与 c460 的相关性

Fig.6 Correlation of images from other 9 cameras with c460

在实验中 $r_{j,j}$ 的分布分别采用对数正态分布、 Γ 分布、 β 分布描述, $r_{j,B}$ 的分布采用 GGD 描述。根据最小错误率的 Bayes 决策规则求出判别阈值 t , 求最小错误率的方法如下

$$S_{\min} = \int_{-1}^t f(r_{j,j})P(j)dr + \int_t^1 f(r_{j,B})P(B)dr, \tag{17}$$

在式(17)中, S_{\min} 表示最小错误率, $P(j)$ 表示来自 j 相机图像样本的先验概率, $P(B)$ 表示其它 9 种相机图像样本的总体先验概率。实验结果如表 1 所示。

表 1 3 台相机的判别阈值 t 和最小错误率 S_{\min}
Tab.1 t and S_{\min} of three cameras

camera	PDF	t	S_{\min}
C5060	log	0.009 2	2.283 6 × 10 ⁻⁷
	gamma	0.007 6	7.357 3 × 10 ⁻⁵
	beta	0.007 4	1.338 3 × 10 ⁻⁴
C460	log	0.010 0	9.156 1 × 10 ⁻⁴
	gamma	0.009 3	0.002 6
	beta	0.009 3	0.002 7
Canon 400	log	0.008 0	0.002 3
	gamma	0.007 6	0.006 0
	beta	0.007 6	0.006 2

由表 1 可知, 采用 3 个不同的 PDF 所得到的判别阈值 t 和最小错误率 S_{\min} 也不同, Γ 分布和 β

分布对应的结果很接近, 但对数正态分布对应的结果与它们相差较大, 且对数正态分布对应的 S_{\min} 最小, 这也说明了采用对数正态分布描述噪声相关性偏态分布的效果最佳。这一结论由图 4 的曲线结合最小错误率的 Bayes 决策规则可以很容易地解释。

比较表 1 和文献[1]中表 2 的数据可知, 采用 Γ 分布、 β 分布对应的结果与选用广义 chi 平方分布对应的结果很接近, 对这 3 种相机而言, 采用对数正态分布(表 1 中的 log)的 S_{\min} 与采用其它 3 种分布的 S_{\min} 相比均降低了 60% 以上, 大幅度地提高了鉴别数字图像真伪的正确性, 这也说明了对噪声相关性偏态分布采用正确的模型是降低鉴别错误率的有效途径。

6 结 论

本文从 PDF 曲线的特征和最小错误率的大小两个角度说明了采用对数正态分布描述噪声相关性偏态分布的效果最佳, 且明显优于文献[1]所介绍的利用广义 chi 平方分布的效果, 更优于文献[3-5]介绍的对噪声相关性的分布仅利用 GGD 描述的效果, 因此本文提出采用对数正态分布描述噪声相关性偏态分布的方法。该方法编程简单、精度高、适应性强, 有较高的应用价值。

由于数码相机使用的各种成像传感器(CCD、CMOS、JFET、CMOS-Foveon X3 等)都是利用半导体制造的, 且制造工艺相似, 它们的模式噪声有相似的特性^[3], 因此本文的方法适合各种成像传感器。由于模式噪声对应着成像传感器像素间的差异, 它就像数码相机的指纹特征, 在成像过程中被嵌入图像信息中^[9], 所以根据噪声相关性的检测原理可知, 相机信噪比的高低不影响本文方法的适用性。

本文介绍的方法没有利用随机变量的几何平均值、 n 阶原点矩、 n 阶中心矩等特征量。因此, 利用有限样本的更多信息去探寻或构造一个能更精确描述噪声相关性偏态分布的 PDF 可能成为下一步工作的研究方向。

参考文献:

[1] 崔夏荣, 苏光大. 数字图像与数码相机噪声相关性

的分布[J]. 中国图象图形学报, 2008, 13(6): 1071-1076.

CUI X R, SU G D. The distribution of noise corre-

- lation between digital image and digital camera[J]. *Journal of Image and Graphics*, 2008, 13 (6): 1071-1076. (in Chinese)
- [2] 崔夏荣. JPEG 压缩对噪声相关性影响的修正[J]. *光电子·激光*, 2009, 20(9):1286-1390.
CUI X R. Modifying the effect of JPEG compression on noise correlation[J]. *Journal of Optoelectronic s · Laser*, 2009, 20(9):1286-1390. (in Chinese)
- [3] LUKAS J, FRIDRICH J, GOLJAN M. Digital camera identification from sensor pattern noise [J]. *IEEE Transactions on Information Foresics and Security*, 2006, 1:205-214.
- [4] CHEN M, FRIDRICH J, GOLIAN M. Digital imaging sensor identification (further study) [J]. *SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX, San Jose, California*, 2007, 6505:16-29.
- [5] FRIDRICH J, CHEN M, GOLJAN M. Imaging sensor noise as digital X-ray for revealing forgeries [C]. *Proc. 9th Int. Workshop on Information Hiding, SantMalo, France*, 2007:342-358.
- [6] 胡琛, 王彬. 基于最大熵原理的分布模型[J]. *山东理工大学学报(自然科学版)*, 2007, 21(6):87-90.
HU CH, WAN B. Distribution model based on maximum entropy principle[J]. *Journal of Shandong University of Technology(Natural Science Edition)*, 2007, 21(6):87-90. (in Chinese)
- [7] 余海, 陈勇, 周钦. 基于最大熵原理的风险分析先验概率分布确定方法[J]. *中国科技信息*, 2008(4): 276-277.
- YU H, CHEN Y, ZHOU Q. The method of ascertaining the prior probability distribution for the risk analysis based on maximum entropy principle [J]. *China Science and Technology Information*, 2008 (4):276-277. (in Chinese)
- [8] 张学文. 组成论[M]. 合肥:中国科技大学出版社, 2003:184-200.
ZHANG X W. *The Constitution Theory* [M]. Hefei: University of Science and Technology of China Press, 2003:184-200. (in Chinese)
- [9] 崔夏荣, 苏光大. 基于噪声相关性的数字图像区域作伪检测[J]. *光子学报*, 2008, 37(10):2108-2113.
CUI X R, SU G D. Detecting spurious area of digital iamge based on noise correlation[J]. *Acta photonica sinica*, 2008, 37(10): 2108-2113. (in Chinese)
- [10] 盛骤, 谢式千. 概率论与数理统计[M]. 北京:高等教育出版社, 2001:436-438.
SHENG ZH, XIE SH Q. *The Theory of Probability and the Mathematical Statistics* [M]. Beijing: Higher Education Press, 2001: 436-438. (in Chinese)
- [11] 罗鹏飞, 张文明. 统计信号处理基础[M]. 北京:电子工业出版社, 2006:473-479.
LUO P F, ZHANG W M. *Translation, Fundamentals of Statistical Signal Processing* [M]. Beijing: Publishing House of Electronics Industry, 2006:473-479. (in Chinese)

作者简介:



崔夏荣(1965—),男,福建浦城人,副教授,1986年于福建师范大学获得学士学位,2005年于武汉理工大学获得硕士学位,主要从事图像处理及小波分析的研究。E-mail:npcxr@163.com