基于混沌神经网络的伪随机序列性能分析*)

陈 军[1] 韦鹏程[1,2] 张 伟[1,2] 杨华千[1,2]
(重庆教育学院计算机与现代教育技术系 重庆 400067)[1]
(重庆大学计算机科学与工程学院 重庆 400044)[2]

摘 要 伪随机序列在保密通信、航空航天、测距、密码学、自动控制等领域具有重要作用。本文结合神经网络和混沌映射的特点,提出了一种基于混沌神经网络和混沌映射混沌伪随机序列的设计方法,该方法可以克服有限精度效应对混沌系统的影响,从而改善混沌序列特性,用理论与计算机仿真实验相结合的方法对混沌序列的随机性、平衡性、相关性和线性复杂度等特性进行了系统的分析。分析结果表明,基于混沌神经网络和混沌映射的混沌伪随机序列具有十分理想的随机特性和相关特性,为在低成本下得到比较实用的序列密码提供了一种新的思路。

关键词 混沌神经网络,混沌映射,伪随机序列

Research on the Performance of Pseudo-Random Sequences Based on Chaotic Neural Network

CHEN Jun[1] WEI Peng-Cheng[1,2] ZHANG Wei[1,2] YANG Hua-Qian[1,2]
(Department of Computer and Modern Education Technology, Chongqing Education College, Chongqing 400067)[1]
(Department of Computer Science and Engineering, Chongqing University, Chongqing 400044)

Abstract Pseudo-Random Sequences play an important role in many fields such as aresecurity communication , aviation , autocontrol and cryptography. in this paper, In order to overcome short period resulted from the finite precision in practical application, combining the characteristic of chaotic neural network and chaotic map, pseudo-random sequences generator base on chaotic neural network and chaotic map is presented. The performances of chaotic sequences are analyzed theoretically, and a lot of analyses and comparison experimentations have been done, including several properties such as randomicity, balance, run length, correlation function and linear complexity. The result shows that this pseudo-random sequence owns ideal.

Keywords Chaotic neural network, Chaotic map, Pseudo-random sequences

1 引言

伪随机序列[1]是由确定性过程产生的,具有类似于白噪声的许多特性,广泛用于如扩频通信、航空航天、测距、密码学、自动控制等领域。目前用的比较多的是由线性反馈移位寄存器(LFSR)生成的序列,如 m-序列,Gold 序列等。但它们普遍存在以下不足[2]:m-序列虽然具有尖锐的自相关特性,但它的互相关特性有较大旁瓣,可用的优码数量少,难以满足 CDMA 对大容量的需求,并且线性复杂度低,容易破译;Gold 序列虽然具有较好的相关特性,但其随机性差。

混沌神经网络模型最早是根据生物神经元的混沌特性于 20 世纪 90 年代初 K. Aihara, T. Takabe 和 M. Toyoda 等人首次提出来的[3,4],它具有非常丰富和复杂的非线性动力学特性,特别是它的混沌动力学特性,它不仅能产生无法预测的伪随机序列列轨迹,而且是一个非常复杂难解的 NP 问题;与以移位寄存器为基础的序列加密法相比,混沌神经网络在序列周期、随机统计性以及线性复杂度方面均有优势,因而这种混沌加密算法的安全性比以移位寄存器为基础的序列加密算法的要高。

利用混沌映射产生混沌序列的理论研究已经很成熟。但是,混沌序列发生器总是在有限精度下实现,混沌迭代过程必将退化为周期序列。本文结合混沌神经网络和混沌映射的特点,提出了一种基于混沌神经网络和混沌映射混沌伪随机序列的设计方法,该方法可以克服有限精度效应对混沌系统的影响,从而改善混沌序列特性。同时,对这类混合混沌序列的周期、平衡、相关性以及线性复杂等特性进行了系统的分析,结果表明,这种混合混沌序列具有随机性好、实现容易、周期长等优点,为在低成本下得到比较实用的序列密码提供了一种新的思路。

2 伪随机序列相关概念

设有二进制序列$\{b_i\}$,其中 $b_i \in \{0,1\}$,如果存在正整数 T,使得对于 $\forall i, b_{i+T}=b_i$,则称$\{b_i\}$为周期序列,满足上述关系的最小 T 称为序列的周期。

若序列$\{b_i\}$,除开始若干项后的其余部分是周期序列,则此序列成为准周期序列。

定义 1[1] 在序列$\{b_i\}$的一个周期中,若

$$b_{i-1}\neq k_i=b_{i+1}=\cdots=b_{i+l-1}\neq b_{i+l} \tag{1}$$

则称$(b_i, b_{i+1}, \cdots, b_{i+l-1})$为序列的一个长为 l 的游程。

定义 2[1] 周期为 T 的序列$\{b_i\}$的周期自相关函数定义为

$$R(j)=\frac{A-D}{T} \tag{2}$$

*)基金项目:重庆市科委自然科学基金资助项目(CSTC,2005B2286),重庆市教委资助项目(No. kj051501),重庆教育学院重点项目。**陈 军** 硕士研究生,主要研究方向为信息安全;**韦鹏程** 博士研究生,主要研究方向为信息安全,混沌理论;**张 伟** 教授,博士后,主要研究方向为信息安全、计算智能与数据挖掘;**杨华千** 博士研究生,主要研究方向为信息安全,混沌数字水印。

式中，$A=|\{0\leqslant i<T;b_i=b_{i+j}\}|$，$D=|\{0\leqslant i<T;b_i\neq b_{i+j}\}|$，$A$ 表示序列$\{b_i\}$和$\{b_{i+j}\}$中相同的位的数目，D 表示序列$\{b_i\}$和$\{b_{i+j}\}$中不同的位的数目。当 j 为 T 的倍数时，$R(j)$为自相关函数，$R(j)=1$；当 j 不是 T 的倍数时，$R(j)$为异相自相关函数。

周期为 T 的伪随机二进制序列应满足 Golomb 提出的三条随机性公设[1]：

① 若 T 为奇数，则序列$\{b_i\}$一个周期内 0 的个数和 1 个数相差 1；若 T 为偶数，则 0 的个数和 1 的个数相等。

② 长度为 T 的周期内，1 游程的个数占游程总数的 1/2，2 游程的个数占游程总数的 $1/2^2$，…，d 游程的个数占游程总数的 $1/2^d$，而任意长度的 0 的游程个数与 1 的游程个数相同。

③ 序列的异相自相关函数 $R(j)$是一个常数。

公设①和②的意义很明确，主要用于衡量序列的平衡性和随机性，而公设③意味着对序列与其平移后的序列作比较，不能获取其它任何信息。满足以上三个条件的序列成为伪随机序列（Pseudo Random Sequence），也称为伪噪声序列（Pseudo Noise Sequence，PN 序列）。

密钥序列往往是个准周期的伪随机序列，但为了满足密码体制的要求，密钥序列除了满足上面的三个条件外，还要符合以下三个要求。

① 期 T 要足够大，如大于 10^{50}；

② 列的产生易于高速生成；

③ 序列的任何部分暴露时，要分析整个序列，提取产生它的电路结构信息在计算上是不可行的，即序列的线性不可预测充分大。

其中的第三点要求确定了密码的强度，是序列密码的核心，它包含了序列密码要研究的许多主要问题，如线性复杂性、相关免疫性、不可预测性等。

3 细胞神经网络和 Chebyshev 映射

细胞神经网络（CNN）理论及其应用是由 Chua 等于 1988 年首先提出的，由于其规则的结构和局部的连接性质而易于超大规模集成电路（VLSI）实现，故 CNN 具有广泛的应用前景。目前，CNN 作为一种灵活而有效的神经网络模型在保密通信、图像处理、模式识别和物理学等领域的应用得到很多学者的关注[5,6]。CNN 的应用在很大程度上取决于其动力学行为，如在图像处理、模式识别和控制中的应用往往需要网络收敛于稳定的平衡点，在保密通信和物理学中的应用往往需要网络具有混沌吸引子或极限环解[1]。本文研究如下三阶 CNN 动态模型：

$$\frac{dx_j}{dt}=-x_j+a_jy_j+\sum_{k=1,k\neq j}^{3}a_{jk}y_k+\sum_{k=1}^{c}S_{jk}x_k+i_j\quad j=1,2,3 \tag{3}$$

这里 x_j 是状态便利，y_j 是相应的输出，满足如下公式：

$$y_j=0.5(|x_j+1|-|x_j-1|)\quad j=1,2,3 \tag{4}$$

如果令：

$a_{12}=a_{13}=a_2=a_{23}=a_{32}=a_3=a_{21}=a_{31}=0$；$S_{13}=S_{31}=S_{22}=0$；$i_1=i_2=i_3=0$；$S_{21}=S_{23}=1$

则系统(3)变成：

$$\begin{cases}\dfrac{dx_1}{dt}=-x_1+a_1y_1+S_{11}x_1+S_{12}x\\[2ex]\dfrac{dx_2}{dt}=-x_2+x_1+x_3\\[2ex]\dfrac{dx_3}{dt}=-x_3+S_{32}x_2+x_3\end{cases} \tag{5}$$

从图 1 可以看出，只要调节非线性扰动参数 a_1、S_{11}、S_{12} 和 S_{32}，系统的运动轨迹变为不同的混沌吸引子。

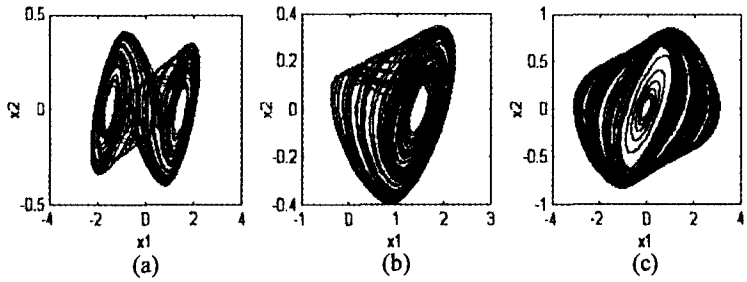

图 1 三阶细胞神经网络的混沌吸引子

(a)$a_1=3.86$，$s_{11}=-1.55$，$s_{12}=8.98$，$s_{32}=-14.25$；(b)$a_1=3.85$，$s_{11}=-1.55$，$s_{12}=8.76$，$s_{32}=-14.35$；(c)$a_1=-4.198$，$s_{11}=2.365$，$s12=7.45$，$s_{32}=-10.98$

Chebyshev 映射(6)是一个典型的混沌系统，其满足绝对连续不变测度、等分布和对称特性的条件。

$$x_{n+1}=\cos(k\arccos(x_n)),-1\leqslant x_n\leqslant 1,n=1,2,3\cdots \tag{6}$$

这里我们主要讨论 $k\geqslant 2k=4$ 的混沌特性。图 2 是两个初始值相差仅为 10^{-5} 的两个混沌时间序列，这表明 Chebyshev 映射具有良好的初始值明性，能产生良好性能的伪随机序列。

4 伪随机序列发生器设计

从理论上讲，对任意给定系统(3)和(6)的初始值，通过迭代会分别产生一个非周期的无穷数值序列，从而对应一个无穷随机序列。该随机序列也是非周期的和类随机的，但计算机精度有限使得实际产生的序列必然具有周期性[7,8]。事实上，我们无法消除周期性，而只能设法延长序列的周期。为此我们提出一种新的思想，用三阶细胞混沌神经网络系统和 Chebyshev 混沌映射来产生伪随机序列，图 2 是其结构图。

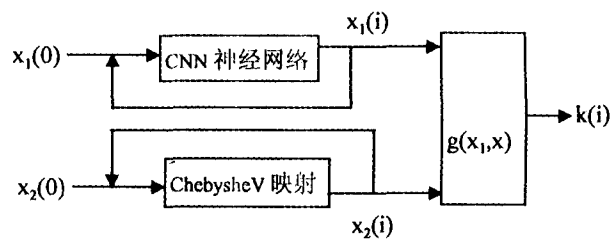

图 2 混沌伪随机序列发生器结构图

伪随机比特序列定义如下：

$$k(i)=g(x_1(i),x_2(i))=\begin{cases}1, & x_1(i)>x_2(i)\\ \text{不输出}, & x_1(i)=x_2(i)\\ 0, & x_1(i)<x_2(i)\end{cases} \quad (7)$$

5 伪随机序列性能分析

基于混沌神经网络和混沌映射伪随机序列满足 Golomb 提出的三条随机性公设。本节我们主要利用计算机仿真实验来考察基于混沌神经网络和混沌映射的伪随机序列的性能。

5.1 0-1 平衡性检验

基于混沌神经网络和混沌映射的伪随机序列随机性好，在整个状态空间服从均匀分布。因此，量化后的二值混沌序列具有理想的统计特性。我们取不同的序列长度，其序列中“1”与“0”的数目几乎趋于平衡(见表 1)。

对实验数据进行自由度为 1 的 χ^2 检验，显著性水平为 5%，对应的 χ^2 的值为 3.761。构造统计量

$$\chi^2=\frac{(n_0-n_1)^2}{N} \quad (8)$$

式中 n_i 表示序列取值 i 的个数。如果该统计量的值小于 3.761，则该序列通过检验。对 100 组长度均为 $N=10000$ 的混沌序列(每组序列对应不同的初值)进行检验，结果该混沌序列通过率为 99%，实验表明基于混沌神经网络和混沌映射的伪随机序列具有很好的平衡性，与理论相符。

表 1 混沌伪随机序列平衡性实验

序列长度 N	5000	10000	20000	30000	40000	50000	100000
0 的个数	2492	4900	9931	14973	19982	25005	50004
1 的个数	2508	5100	10069	15027	20018	24995	49996
不平衡度	0.032%	0.020%	0.0069%	0.0018%	0.0009%	−0.0002%	−0.00008%

5.2 序列检验

序列检验用来判定转移概率是否合理，即出现相同和不相同相邻元素的概率大致相等。令 n_{00} 表示 00 的个数，n_{11} 表示 11 的个数，n_{10} 表示 10 的个数，n_{01} 表示 01 的个数。取统计量为：

$$\chi^2=\frac{4}{n-1}\sum_{i=1}^{1}\sum_{j=0}^{1}(n_{ij})^2-\frac{2}{n}\sum_{i=0}^{1}(n_i)^2+1 \quad (9)$$

对于自由度为 2 的 χ^2 分布，可得到对应的 5% 显著性水平的 χ^2 值是 5.985。对 100 组长度均为 $N=10000$ 的混沌序列(每组序列对应不同的初值)进行检验，通过率均为 98.5%。

5.3 游程特性

游程特性即 Golomb 提出的序列随机性公设的第 2 条。我们对混沌序列分别取不同值时进行了游程特性实验，并与级数为 17 的 m-序列的游程特性进行了比较，结果见表 2。从表中看出，基于混沌神经网络和混沌映射的伪随机序列的游程特性优于 m-序列的游程特性。

5.4 相关特性

定义 3[9] 设 $a^{(1)}$，$a^{(2)}$ 分别表示长度为 N 的两个不同混沌伪随机序列。序列 $a^{(1)}$ 和 $a^{(2)}$ 的非周期互相关函数为：

$$C_{a^{(1)},a^{(2)}}(l)\begin{cases}\frac{1}{N}\sum_{i=0}^{N-1-l}a_i^{(1)}(a_{i+l}^{(2)})^*, & 0\leqslant l<N\\ \frac{1}{N}\sum_{i=0}^{N-1+l}a_{i-l}^{(1)}(a_i^{(2)})^*, & 1-N\leqslant l<0\\ 0, & |l|\geqslant N\end{cases} \quad (10)$$

当 $a^{(1)}=a^{(2)}$ 时，上式表示的函数是扩频序列的非周期自相关函数。

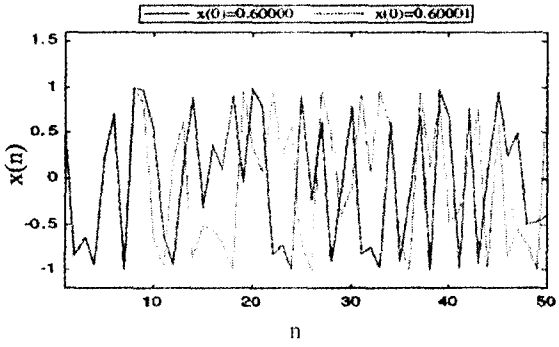

图 3 初始值相差 10^{-5} 的 Chebyshev 时间序列

我们对基于混沌神经网络和混沌映射的伪随机序列进行相关特性检测，取序列长度为 2000，相关间隔为 −500～500，其非周期自相关与互相关特性如图 4，由实验结果看出，混沌序列具有尖锐的自相关特性和很小的互相关值。

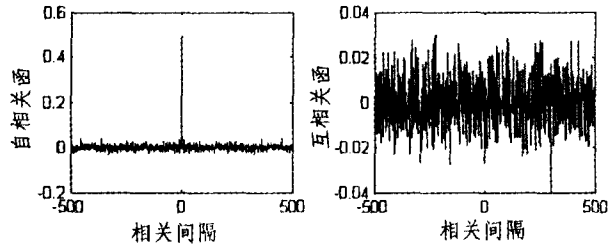

图 4 混沌伪随机序列相关函数

5.5 线性复杂度

从理论上获得混沌序列的线性复杂度的表达式目前尚有困难。此处我们应用 Berlekamp-Massey 算法来计算序列的线性复杂度[10]。结果如表 3 所示。从表中我们得知基于混沌神经网络和混沌映射的混沌伪随机序列具有理想的线性复杂度，即 $L\approx N/2$。

表 2 混沌伪随机序列游程特性

分项统计		序列长度 N			
		128	256	1024	2048
混沌伪随机序列	1 游程	0.2525	0.2508	0.2503	0.2492
	2 游程	0.2540	0.2526	0.2513	0.2503
	3 游程	0.1939	0.1941	0.1947	0.1939
	4 游程	0.1249	0.1258	0.1256	0.1251
	短游程和	0.8253	0.8233	0.8220	0.8185
m-序列	1 游程	0.2544	0.2526	0.2510	0.2508
	2 游程	0.2568	0.2556	0.2547	0.2544
	3 游程	0.1806	0.1806	0.1808	0.1812
	4 游程	0.1267	0.1275	0.1276	0.1282
	短游程和	0.8185	0.8163	0.8141	0.8146

表 3 混沌序列的线性复杂度

线性复杂度	序列长度 N						
	256	512	1024	2048	4096	8192	16384
混沌伪随机序列	127	255	511	1024	2048	4095	8192

(下转第 177 页)

度值升序排列后的散点图。

表 1 孤立点挖掘结果列表

name	X_1	X_2	X_3	X_4	density
ST 博讯	−0.015	0.08	0.08	−19	9.169E-7
千金药业	0.19	11.18	11.45	1.67	5.652E-4
伊泰 B 股	0.38	3.41	3.67	10.3	1.147E-3
济南钢铁	0.39	4.05	4.05	9.58	1.393E-3
恒源煤电	0.38	6.78	6.78	5.6	2.122E-3
马应龙	0.29	10.59	10.61	2.71	2.269E-3
烟台万华	0.18	2.22	2.22	10.82	4.260E-3
云天化	0.3136	4.36	4.4	7.13	4.757E-3

表 2 四个财务指标的平均值与标准差

	X_1	X_2	X_3	X_4
平均值	0.08008	3.22628	3.29674	2.30820
标准差	0.08589	1.60271	1.58628	2.93456

表 3 各指标相关系数

	X_1	X_2	X_3	X_4
X_1	1.00000	0.46550	0.46405	0.72010
X_2	0.46550	1.00000	0.99493	0.10048
X_3	0.46405	0.99493	1.00000	0.09949
X_4	0.72010	0.10048	0.09949	1.00000

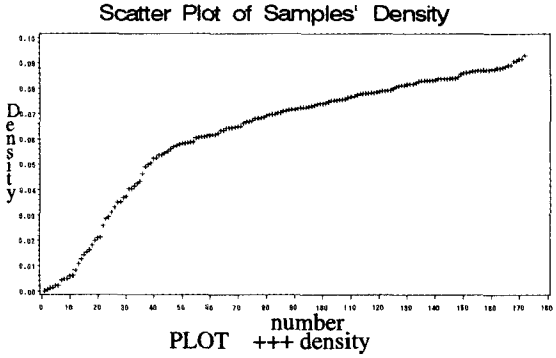

图 2 样本独立成分联合密度的散点分布图

从表 3 可以看出，X_2 与 X_3 具有很大的相关性，根据指标筛选方法，我们最后选择指标 X_1、X_3 和 X_4 进行独立成分分析。从表 1、表 2 中可以看出，第一个孤立点“ST 博讯”指标 X_4 明显偏离中心值，第二个孤立点“千金药业”的指标 X_1，X_4 明显偏离中心值，但偏离程度不如第一个大，找出的前面 8 个孤立点至少都有一个指标明显偏离中心值，与实际情况基本吻合。从图 2 我们也可以看出，样本大部分集中在密度值比较大的地方。需要说明的是，孤立点挖掘的结果，只是给用户提供一个参考，只有用户才能最后确定真正的孤立点。

结论 挖掘孤立点，发现有价值的隐藏信息，是数据挖掘的一项重要内容。传统的孤立点挖掘方法实际上是假定数据之间是互相独立的，或事先假定数据服从高斯分布，但实际生活中很多高维数据之间存在一定的相关性，且很多数据不服从高斯分布，这些缺点限制了传统孤立点挖掘方法的应用。本文提出的 ISOM 模型，先用 ICA 对数据进行独立成分分解，再用 SVM 估计各独立成分的密度函数，克服了传统孤立点挖掘方法的局限性，为数据挖掘提供了一种有效的方法。实验结果也验证了本文提出的 ISOM 模型的有效性与合理性。

参 考 文 献

1 Liu Xiao-Hui. Strategies for outlier analysis. Birkbeck College University of London, 2000
2 Edwin M K, Raymond T Ng. Algorithm for Mining Distance-Based Outliers in Large Databases. In: Proc. of the 24th VLDB Conf. New York, USA, 1998
3 Johanna H, Rocke D M. Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator. Computational Statistics & Data Analysis, 2004, 44: 625～638
4 Ester M, et al. A Density-Based Algorithm for Discovering Clusters in large spatial databases. In: Proc. of 2nd Intl. Conf. on Knowledge Discovery and Data Mining, 1996
5 Bayarri M J, Morales J. Bayesian measures of surprise for outlier detection. Journal of Statistical Planning and Inference 2003, 111: 3～22
6 Bullen R J, et al. Outlier detection in scatterometer data: neural network approaches. Neural Networks (in press)
7 Kantardzic M. Data Mining Concepts, Models, Methods, and Algorithms. Tsinghua University Press, 2003
8 De Groot P J, Postma G J, et al. Application of principal component analysis to detect outliers and spectral deviations in near-field surface-enhanced Raman spectra. Analytica Chimica Acta, 2001, 446: 71～83
9 Jutten C, Herault J. Independent component analysis versus PCA. In: Proc. of European Signal Processing Conf. 1988. 287～314
10 Yogesh S. A simplified approach to independent compoent analysis. Neural Comput & Applic. 2003, 12: 173～177
11 Kocsor A, Csirik J. Fast Independent Component Analysis in Kernel Feature Spaces. LNCS, 2001, 2234: 271～281
12 Cotes C, Vapnik V. Support Vector networks. Machine Leaning, 1995, 20: 273～295
13 Bartlett P L, Shawe-Taylor J. Generalization performance on support vector machines and other pattern classifiers. In: B. Sholkopf, C. Burges, A. Smola, eds. Advances in Kernel Methods-Support Vector Learning, Cambridge, MA: MIT Press, 1999
14 Vapnik V N. Statistical Learning Theory. Publishing House of Electronics Industry, 2004
15 彭红毅，朱思铭，蒋春福. 数据挖掘中基于 ICA 的缺失数据值的估计. 计算机科学，2005，32(12)：203～205

(上接第 163 页)

总结 本文重点研究了基于混沌神经网络和混沌映射的混沌伪随机序列的性能。用理论与计算机仿真实验相结合的方法对混沌序列的随机性、平衡性、相关性和线性复杂度等特性进行了系统的分析，并给出了相应的特性曲线。分析结果表明，基于混沌神经网络和混沌映射的混沌伪随机序列具有十分理想的随机特性和相关特性。由于混沌序列的产生非常方便，数量众多，因此可以用来替代 m-序列，以满足 CDMA 通信对大容量的需求。混沌序列的线性复杂度高，不容易破译，因此除了用于扩频通信之外，还可以作为传统密码学中的密钥来使用。

参 考 文 献

1 肖国镇著. 伪随机序列及其应用. 北京：国防工业出版社
2 Diffie W, Hellman M E. New directions in cryptography. IEEE Trans. on Information Theory, 1976, 22 (6): 644～654
3 van Schyndel R G, Tirkel A Z, Svalbe I D. Key independent watermark detection. IEEE International Conference on Multimedia Computing and Systems, Florence, Italy, 1999. 580～585
4 Eggers J J, Su Jonathan K, Girod B. Public key watermarking by eigenvectors of linear transforms. In: European Signal Processing Conference, Tampere, Finland, 2000. 428～435
5 Chua L O, Roska T. The CNN paradigm. IEEE Trans. CAS-I, 1993, 40: 47～156
6 Civalleri P P, Gilli M. On dynamic behaviour of two-cell cellular neural networks, Int. J. Circ. Th. Appl., 1993, 21: 451～471
7 丘水生，陈艳峰，吴敏，等. 混沌加密的若干问题与新的加密系统方案. 见：2002 中国非线性电路与系统学术会议论文集. 中国：深圳，2002，11：174～179
8 王育民. 混沌密码序列使用化问题. 西安电子科技大学学报，1997，24(4)：560～562
9 Tohur K, Akio T. Pseudonoise sequence by chaotic nonlinear and their correlation properties. IEICE Trans commun, 1993, E97-B (8): 855～862
10 Rueppel R A. Linear complexity and random sequences, Advances in Cryptology. EURO CRYPT'85 (LNCS 219), 1986. 167～188
11 何振亚，张毅锋，卢宏涛. 细胞神经网络动态特性及其在保密. 通信中的应用. 通信学报，1999，20(3)：59～67