基于协同神经网络的图像数字水印算法

陈永强[1),2)] 胡汉平[1)] 李新天[1)]

[1)](华中科技大学图像识别与人工智能研究所,图像信息处理与智能控制教育部重点实验室,武汉 430074)

[2)](武汉工业学院计算机与信息工程系,武汉 430023)

摘 要 为保证图像数字水印的鲁棒性和安全性,结合加密技术,提出了一种使用协同神经网络的图像数字水印算法。该算法是先对有意义的灰度图像进行水印序列化处理后,再将其作为水印信号嵌入到载体图像的分块 DCT 系数直流分量里。水印的检测提取算法是采用协同神经网络,将疑似水印信号作为网络的输入,而网络的输出就是识别的结果。经仿真实验验证,该算法可承受一定的图像处理操作,不仅能同时完成水印的检测与提取,而且能有效判断水印的归属,具有较好的性能。

关键词 数字水印 协同学 神经网络 数字图像处理

中图法分类号:TP309.2 TP391.4 **文献标识码**: A **文章编号**: 1006-8961(2005)07-0894-06

Image Digital Watermark Algorithm Based on Synergetic Neural Network

CHEN Yong-qiang[1),2)], HU Han-ping[1)], LI Xin-tian[1)]

[1)](*Institute for Pattern Recognition and Artificial Intelligence, State Education Department Key Laboratory for Image Processing and Intelligent Control, Huazhong University of Science and Technology, Wuhan* 430074)

[2)](*Department of Computer and Information, Wuhan Polytechnic University, Wuhan* 430023)

Abstract An digital image watermark algorithm is proposed based on the synergetic neural network, combined with encryption technique, to guarantee its robustness and security. The meaningful gray image is serially processed into watermark signal and embedded into the DC elements of the block DCT coefficient matrix of the host image. Watermark detection/extraction algorithm is realized by using synergetic neural network. The network's input is possible watermark signal and its output is the recognition result. The emulational experiments indicates that, the proposed algorithm can fulfill the watermark detection/extraction at one time after image processing, and identify the watermark owner which shows good performance.

Keywords digital watermark, synergetics, neural network, digital image process

1 引 言

随着信息技术和计算机网络的飞速发展,数字式多媒体信息(图像、音频、视频、文本)的存储、复制与传播变得非常方便,但同时对多媒体信息的侵权也不断发生,因此对多媒体内容的版权保护已成为亟待解决的问题。由于现有的版权保护系统和传统的加密方法对多媒体内容的保护和完整性认证方面具有一定的局限性,因而数字水印技术随之提出。数字水印技术是将数字、序列号、文字、图像标志等版权信息嵌入到多媒体数据中,由于其嵌入的信息能量低,不会被人察觉,可起到版权保护、秘密通信、数据文件的真伪鉴别和产品标识等作用,因而在多媒体信息的版权保护与完整性认证方面得到了广泛应用。

众所周知,数字水印应该具有可证明性、不可感知性、鲁棒性和安全性等基本特性[1]。其中鲁棒性和安全性的特性要求数字水印在经过无意、善意的

收稿日期:2004-10-22;**改回日期**:2005-01-12

第一作者简介:陈永强(1967~),男,讲师。2001 年获华中科技大学机械设计与理论专业硕士学位,现为该校模式识别与智能系统博士研究生。主要研究方向:网络信息安全与智能系统,数字图像处理。E-mail: chenyqwh@hotmail.com

常规处理和恶意攻击后,具有能够检测或提取到水印的能力。目前主要采用扩频通信和密码学等技术来实现相应的鲁棒性和安全性要求。Cox 等人应用扩频技术,使用一个随机向量,通过改变一幅图像中1 000个感知上最重要的离散余弦变换(DCT)系数来实现水印的嵌入,证明有较好的鲁棒性[2]。为防止未经授权水印的嵌入和检测,Cox 等人和 Pitas 分别提出了私有和公有水印方案[3,4]。Craver 等人在分析私有和公有水印不能防止未经授权水印删除缺点的基础上,设计了具有非可逆性的单向水印方案[5],但其不能抵抗伪造水印图像的攻击。

在数字水印里,神经网络主要应用于水印嵌入时的图像分类或产生自适应于图像的水印[6]以及水印的检测[7]等两个方面,实验表明,其对水印的鲁棒性和安全性性能有较好的提高作用。本文由此出发,提出了一种运用协同神经网络的 DCT 域图像数字水印检测提取算法。虽然协同神经网络在人脸、车牌等模式识别领域获得了较多应用,但在数字水印里还未见报道。

2　协同神经网络模型

协同学是研究系统依靠自组织产生空间结构、时间结构或功能结构上自发形态的一门跨学科的研究领域,其研究的焦点是复杂系统宏观特征的质变,以及如何描述系统在演化过程中的宏观有序。协同神经网络是用协同学理论所构造的,其与传统的从研究单个神经元的特性、配置和连接的构造方法完全不同,它是一种自上而下的神经网络[8]。

2.1　动力学演化方程

一个协同联想模式识别系统可用动力学演化过程来描述,其能反映识别系统通过神经网络学习演化来补全不完全的数据集合和模式形成的过程。假设原型模式数为 M,原型模式向量维数为 N,并要求 $M \leqslant N$,则有以下动力学方程:

$$\boldsymbol{q}' = \sum_k \lambda_k \boldsymbol{v}_k (\boldsymbol{v}_k^+ \boldsymbol{q}) - B \sum_{k \neq \hat{k}} (\boldsymbol{v}_{\hat{k}}^+ \boldsymbol{q})^2 (\boldsymbol{v}_k^+ \boldsymbol{q}) \boldsymbol{v}_k - C(\boldsymbol{q}^+ \boldsymbol{q}) \boldsymbol{q} + \boldsymbol{F}(t) \tag{1}$$

其中,$\boldsymbol{q}$ 是以输入模式 $\boldsymbol{q}(0)$ 为初始值的状态向量,其为待识别的模式向量,可分解为原型向量 $\boldsymbol{v}_k$ 和剩余量 $\boldsymbol{w}$,且 $\boldsymbol{q} = \sum_{k=1}^{M} \xi_k \boldsymbol{v}_k + \boldsymbol{w}$,$\boldsymbol{v}_k^+ \boldsymbol{w} = 0$;$\boldsymbol{q}^+$ 为 $\boldsymbol{q}$ 的伴随向量,λ_k 为注意参数,只有当它为正的时候,模式才能被识别;$\boldsymbol{F}(t)$ 为涨落力,可忽略不记;B 和 C 为指定系数,且都大于 0;$\boldsymbol{v}_k$ 为原型模式向量,$\boldsymbol{v}_k = (v_{k,1}, v_{k,2}, \cdots, v_{k,N})^{\mathrm{T}}$;$\boldsymbol{v}_k^+$ 为 $\boldsymbol{v}_k$ 的伴随向量,且

$$\boldsymbol{v}_k^+ \boldsymbol{v}_k = \delta_{k,\hat{k}} = \begin{cases} 1, & k = \hat{k} \\ 0, & k \neq \hat{k} \end{cases}$$

$\boldsymbol{v}_k$ 必须满足以下归一化和零均值条件:

$$\sum_{l=1}^{N} v_{k,l} = 0,\ \|\boldsymbol{v}_k\|_2 = \left(\sum_{l=1}^{N} v_{k,l}^2 \right)^{1/2} = 1 \tag{2}$$

2.2　协同神经网络

使用序参量 $\xi_k = \boldsymbol{v}_k^+ \boldsymbol{q}$ 和 $D = (B + C) \sum_{\hat{k}} \xi_{\hat{k}}^2$ 来得到用序参量描述的动力学演化方程为

$$\xi_k' = \xi_k (\lambda - D + B\xi_k^2) \tag{3}$$

由此构造的协同神经网络分为 3 层(如图 1 所示),其中上层是输入层,输入层的单元 j 接收待识别模式向量初始值 $\boldsymbol{q}(0)$ 的第 j 个分量 $q_j(0)$;中间层表示各个序参量神经元,序参量 ξ_k 是由每个输入值 $q_j(0)$ 乘以相连接的 $v_{k,j}^+$,并对全部角码 j 求和所得,网络运行时,具有活性的序参量 ξ_k 的各个神经元即可识别出角码 k 所确定的特定原型模式,网络按照动力学方程运行和演化,同时随时间的发展达到终态,并通过 D 的侧抑制相互作用来进行竞争,最后只有一个序参量能幸存,利用该序参量即可得到 q_j;下层是输出层,输出层的模式可表达成 $q_j(t) = \sum_k \xi_k(t) v_{k,j}$,$q_j$ 是输出层单元 j 的活性,ξ_k 是中间层的最终状态。当 $k = k_0$ 时,$\xi_k = 1$;其他情况下,$\xi_k = 0$。$v_{k,j}$ 是原型向量 $\boldsymbol{v}_k$ 的第 j 个分量,另外,可通过用 $u_{k,j}$ 替换向量的分量 $v_{k,j}$ 来识别出用 $u_{k,j}$ 描述的属于角码 k 的新模式。

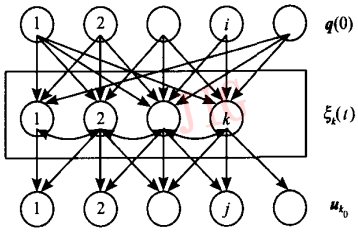


图 1　协同神经网络

Fig. 1　Synergetic neural network

3　图像数字水印嵌入算法

3.1　数字水印通用算法

通用数字水印算法一般包含水印嵌入和水印检

测或提取两个过程,其中水印可由多种模型构成,如随机数字序列、数字标识、文本以及图像等,并常需要对水印进行随机化以及加密处理。

设 $\boldsymbol{c}$ 为原始载体信号,$\boldsymbol{m}$ 为水印信号,$\boldsymbol{K}$ 为密码,水印的嵌入过程为

$$\boldsymbol{c}_{\mathrm{w}} = E_{\mathrm{e}}(\boldsymbol{c}, F(\boldsymbol{c}, \boldsymbol{m}, \boldsymbol{K})) \tag{4}$$

函数 F 是一个加密过程,水印 $\boldsymbol{m}$ 通过函数 F 加密处理后,再运用嵌入函数 E_{e} 在载体信号的时(空)域或变换域中嵌入水印,即形成含水印的载体信号 $\boldsymbol{c}_{\mathrm{w}}$。

水印检测提取过程为

$$\hat{\boldsymbol{m}} = D_{\mathrm{e}}(\hat{\boldsymbol{c}}_{\mathrm{w}}, \boldsymbol{m}, \boldsymbol{c}, \boldsymbol{K}) \tag{5}$$

在水印检测提取过程中,首先是做一个判定水印存在与否的 0-1 决策,再在水印存在的前提下,从接收到的含水印载体信号 $\hat{\boldsymbol{c}}_{\mathrm{w}}$ 中,运用提取函数 D_{e} 来提取,并用密码 $\boldsymbol{K}$ 解密还原出水印。含水印的载体信号 $\boldsymbol{c}_{\mathrm{w}}$ 在传输和使用中,会因受到可能的常规信号处理和攻击而由 $\boldsymbol{c}_{\mathrm{w}}$ 改变为 $\hat{\boldsymbol{c}}_{\mathrm{w}}$。提取水印时,需根据是否使用原始信息水印 $\boldsymbol{m}$ 或原载体信号 $\boldsymbol{c}$ 来确定提取函数 D。与水印的嵌入过程相比,水印的提取是水印算法中最重要的步骤。

3.2 水印序列生成

有意义水印信号是代表一定意义的文本、声音、图像或视频信号。使用有意义的水印,提取出的水印直观,能直接对载体中是否含有水印进行判别。为保证水印的可靠性和安全性,有意义的水印信号应加密为水印信号序列。设有意义的水印灰度图像大小为32×32,其处理为水印序列的过程如下:

(1)对水印灰度图像做 DCT 变换,首先得到 32×32大小的 DCT 实系数矩阵;

(2)将系数矩阵 Zig-Zag 向量化,即 2 维矩阵以 Zig-Zag 顺序读取,转换成具有1 024个元素的 1 维向量;

(3)选取密钥 $\boldsymbol{K}$,用某一加密算法,加密 1 维向量;

(4)对加密后的 1 维向量进行从大到小排序来得到水印信号的分量 $m_i, i=1,2,\cdots,1\,024$,将其作为嵌入的水印信号,并保留排序前后的位置对应关系。

3.3 水印嵌入

载体信号是 256×256 的灰度数字图像,水印信号嵌入到载体图像的 DCT 域中,嵌入的具体方法和步骤如下:

(1)将载体图像像素灰度值矩阵 $\boldsymbol{f}=[f(x,y)]$ 均匀分成互不覆盖的1 024个大小为 8×8 的块 $\boldsymbol{f}_i=[f_i(\hat{x},\hat{y})]$,其中,$0\leqslant x,y\leqslant 255$,$0\leqslant \hat{x},\hat{y}\leqslant 7$,$i=1,2,\cdots,1\,024$;

(2)对每块进行 8×8 的 DCT 变换来得到1 024个 2 维 DCT 系数矩阵 $\boldsymbol{F}_i=[F_i(\hat{u},\hat{v})]$,$0\leqslant \hat{u},\hat{v}\leqslant 7$;

(3)对每个 DCT 系数矩阵的 $F_i(0,0)$ 进行从大到小的排序来得到 $\hat{F}_i(0,0)$,并保留排序前后的位置对应关系;

(4)采用式 $\tilde{F}_i(0,0)=\hat{F}_i(0,0)+\alpha\cdot m_i$,$\alpha$ 为嵌入强度,把水印信号分量 m_i 依次嵌入到各个图像块的 DCT 系数直流分量里,而其他的非直流分量保持不变,由此得到 $\tilde{F}_i(\hat{u},\hat{v})$;

(5)对 DCT 域中调整后的系数矩阵进行 DCT 反变换,再根据排序位置关系还原,其得到的含水印图像的灰度值为 $f_{\mathrm{w}}(x,y)=\bigcup_{i=1}^{1\,024}\mathrm{IDCT}\{\tilde{F}_i(\hat{u},\hat{v})\}$,进而可获得嵌入水印的载体图像 $\boldsymbol{f}_{\mathrm{w}}=[f_{\mathrm{w}}(x,y)]$。

4 图像数字水印检测提取算法

由于含水印的图像载体在传输和使用中,会受到一定程度的处理和攻击,因此在处理和攻击超过一定的强度下,直接检测提取到的有意义水印可能从肉眼上无法清楚地判断。在图像载体中检测提取水印,可认为是在大量的水印模式中形成和识别一个特定存在的水印的过程,而且能使用基于协同神经网络的模式识别方法进行识别。由于协同序参量 ξ_k 代表了原型模式和试验模式之间的一种匹配度量,因此在协同神经网络中,经过学习和训练即可识别所输入水印的模式归属。

检测水印时,首先选择若干个与原始水印图像大小相同、内容相似的灰度图像,加上水印图像,总共 M 幅灰度图像组成原型模式初始集。与水印序列的生成过程一样,可通过对每幅图像进行 DCT 变换、Zig-Zag 向量化、加密和排序来得到向量 $\boldsymbol{v}_k=(v_{k,1},v_{k,2},\cdots,v_{k,N})^{\mathrm{T}}$,$k=1,2,\cdots,M$;$N=1\,024$。只要满足 $M\leqslant N$,M 个模式向量就可组成原型模式向量集。

根据协同神经网络模型,协同神经网络的学习算法是通过原型模式向量求其伴随向量的训练过程,即

(1)依式(2),计算满足归一化和零均值条件的原型模式向量 $\boldsymbol{v}_k$;

(2)计算原型模式向量 $\boldsymbol{v}_k$ 的伴随向量 $\boldsymbol{v}_k^+$。

使用协同神经网络方法,能根据网络输出的识别模式的结果,同时完成水印的检测和提取过程。确定协同动力学方程参数 $B=C=1$ 和 $\lambda_k=\lambda=1$,协

同神经网络的识别过程为:

(1)进行水印信号嵌入的逆过程,计算满足归一化和零均值条件的实验模式向量 $\boldsymbol{q}(0)=\{q_j(0)\}$,$j=1,2,\cdots,N$;

(2)由实验模式向量 $\boldsymbol{q}(0)$ 和伴随向量 $\boldsymbol{v}_k^+$,计算序参量 ξ_k 的初值 $\xi_k(0)$。根据协同支配原理,最大的初始序参量值所代表的模式在协同竞争演化后会获胜,即可初步检测出载体图像中包含有水印;

(3)将 $\xi_k(0)$ 输入到协同神经网络,并进行协同联想动力学演化,系统演化结果的终态取决于输入向量的初始序参量值,即在竞争中,具有最大初始序参量的 $\boldsymbol{v}_k$ 获胜,其序参量 ξ_k 趋向于 1,而其他序参量则趋向于 0,这样网络就识别和恢复出加密水印序列。最后对加密的序列进行解密,即可得到所嵌入的图像水印原图。

5　仿真实验与分析

为验证本文算法效果,从 ORL 人脸库中选择 5 幅人脸正面灰度图像[9]进行了仿真实验,人脸实验图像都缩小为 32×32 大小的图像,组成模式初始集(如图 2 所示)。实验时,采用 DES 对称加密算法分别对这 5 幅图进行处理,首先得到由 5 个 1 维向量组成的原型模式向量集,载体图像使用大小为 256×256 的灰度 Lena 图,设置嵌入强度参数 $\alpha=0.1$;然后把原型模式向量集中的第 1 个向量作为有意义水印嵌入到 Lena 图像里,即得到含水印的载体图像(如图 3(b)所示)。

图 2　模式初始集

Fig. 2　Original pattern set

(a)载体图像　　(b)含水印载体图像

图 3　水印嵌入示意

Fig. 3　Watermark embedding

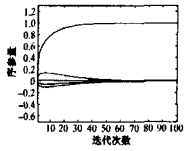
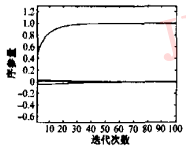
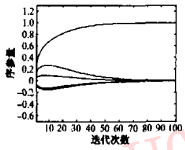
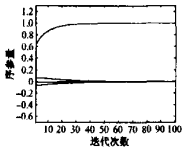
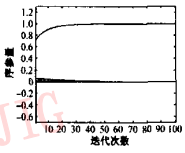
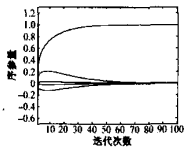
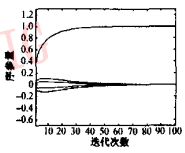
常规图像处理包括加性噪声、幅度变化、线性滤波、有损压缩和几何失真,其目的是为了考察图像水印的鲁棒性。在 MATLAB 环境中,这些处理分别是对含水印 Lena 图像加均值 0,方差从 0.000 1 到 0.009 的高斯噪声、加均值 0,方差从 0.001 到 0.009 的椒盐噪声、像素灰度值范围从[0.1,0.9]到[0,1]的对比度增强、像素灰度值范围从[0,1]到[0.1,0.9]的对比度减弱、3×3 的中值滤波和维纳滤波、压缩质量从 90% 到 10% 的 JPEG 压缩、±1°以内的几何旋转、1.01 到 1.1 倍的几何放大等仿真操作,然后应用上述算法检测提取水印,其所得部分结果如表 1 所示。表中第 1 列是对含水印 Lena 图像所作的图像处理类型和强度;第 2 列是进行相应处理后的含水印 Lena 图;第 3 列是从 Lena 图像中检测到水印后,直接用图像水印嵌入算法的逆过程提取水印,且在不进行协同神经网络演化情况下,经解密得到的水印图像;第 4 列为各序参量在协同神经网络中的动力学协同竞争演化曲线图,其中横坐标为迭代运算的步数,纵坐标为序参量的取值。

从实验结果可知,除幅度变化检测提取不到正确的水印外,对其他的处理类型,协同神经网络在演化运算 100 步以内就可有结果,且嵌入的水印所对应模式的序参量 ξ_k 趋向于 1,即获得了正确的识别结果。从直接提取的水印图像中,不易判断嵌入的水印图像,而采用协同神经网络进行识别,则结果的判断是简单和确定的。仿真实验和与相关系数法比较的结果说明,基于协同神经网络的图像数字水印算法不仅可承受较高强度的高斯噪声、椒盐噪声、JPEG 压缩、中值滤波、维纳滤波、几何旋转和放大等图像处理和几何失真的操作,而且对这些操作具有很强的鲁棒性。

在水印算法公开的前提下,数字水印算法的安全性由密钥的长度所决定。如果已知水印的嵌入和提取算法,那么只要选择足够长的密钥,则该水印算法就是安全的。从协同神经网络的工作原理可知,在原型模式向量集里,由于没有伪造水印的原型模式,再加上伪造水印与原水印的相似程度比不上原水印自身间的相似程度,即原水印的协同序参量大,其在协同竞争演化中会最终获胜,同时协同神经网络只会识别出原水印,而不会识别出伪造水印,即不存在伪状态,因此水印算法能有效地抵抗伪造水印图像的攻击。

表 1 常规图像处理

Tab. 1 General image process

处理方式	处理后含水印图	直接提取	协同神经网络演化
高斯噪声 均值 0,方差 0.009			
椒盐噪声 均值 0,方差 0.009			
JPEG 压缩压缩质量 10%			
中值滤波			
维纳滤波			
旋转 0.5°			
放大 1.05 倍			

6 结 论

用协同学的方法构造神经网络是研究和设计神经网络的一种新颖方法,由于它是一种自上而下的结构构造方法,因此它克服了传统的神经计算中所遇到的学习和模式识别两方面的困难。

水印的检测提取是数字水印算法的关键步骤,如何判断载体图像是否包含水印或水印的版权归属就显得至关重要。图像数字水印的嵌入者可采用水印嵌入算法把有意义版权信息嵌入到图像数据中,在完成自己的图像数字文件的版权标识之后,建立

自己的版权水印模式集,进而采用协同神经网络的水印检测与提取算法来同时确认在他人使用的图像里,是否包含并提取有自己的水印版权信息,以便依此验证水印的归属。

虽然基于协同神经网络的图像数字水印算法不能承受图像的幅度变化,但由于人眼对灰度图像中灰度级层次的改变特别敏感,不宜采用此类图像处理方法,因此文中提出的水印算法对一般的图像处理具有很好的鲁棒性和安全性。

参考文献(References)

1 Cox I J, Miller M L, Bloom J A. Digital watermarking[M]. San Francisco, CA,USA: Morgan Kaufmann Publishers, 2002. [Cox I J, Miller M L, Bloom J A 著. 王颖,黄志蓓译. 数字水印[M]. 北京:电子工业出版社,2003.]

2 Cox I J, Kilian J, Leighton F T, *et al.* A secure robust watermark for multimedia[A]. In: Proceedings of the First International Workshop on Information Hiding[C], Cambridge, UK, 1996:185 ~ 206.

3 Cox I J, Kilian J, Leighton F T, *et al.* Secure spread spectrum watermarking for images, audio and video[A]. In: Proceedings of the 1996 IEEE International Conference in Image Processing Part 3 [C], Lausanne, Switzland, 1996:243 ~ 246.

4 Pitas I. Method for signature casting on digital images[A]. In: Proceedings of the 1996 IEEE International Conference in Image Processing Part 3[C], Lausanne, Switzland, 1996:215 ~ 218.

5 Craver S, Memon N, Yeo L, *et al.* Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks, and implications [J]. IEEE Journal on Selected Areas in Communications, 1998, **16**(4):573 ~ 586.

6 Davis K J, Najarian K. Maximizing strength of digital watermarks using neural networks[A]. In: International Joint Conference on Neural Networks V4[C], Washington, DC, USA, 2001, **4**:2893 ~ 2898.

7 Yu Pao-Ta, Tsai Hung-Hsu, Lin Jyh-Shyan. Digital watermarking based on neural networks for color images[J]. Signal Processing, 2001, **81**(3):663 ~ 671.

8 Hanken H. Synergetic computers and cognition—a top-down approach to neural nets[M]. Berlin: Springer-Verlag, 1991.

9 AT&T Laboratories Cambridge. The ORL database of faces[DB/OL]. http://www.uk.research.att.com/facedatabase.html. 2004-09-01.