

文章编号: 1674-9057(2014)04-0775-07

doi:10.3969/j.issn.1674-9057.2014.04.030

融合灰色马尔科夫理论的二进小波图像的 复制-粘贴篡改检测算法

张瑞芳¹, 程晓辉¹, 宋子航¹, 陈扬²

(1. 桂林理工大学 信息科学与工程学院, 广西 桂林 541004; 2. 桂林市公安局, 广西 桂林 541001)

摘要: 提出了一种二进小波图像的篡改检测算法, 对待测图像进行二进小波图像分解, 然后提取其每个图像分块的高低频段图特征值进行分析。为了使该算法能够在大量图像篡改检测过程中保证其检测率、错误率, 使用基于灰色理论与马尔科夫链的融合模型对该算法的检测率、错误率进行预测, 将预测值进行均值计算得到平均检测率与平均错误率。通过与文献算法比较检测率、错误率均值, 证明了该算法在避免偶然的情况下, 提高了检测率并降低了错误率。

关键词: 二进小波; 复制粘贴篡改; 边缘处理; 灰色-马尔科夫链

中图分类号: TP391.41

文献标志码: A

近年来随着数码产品的迅速普及以及互联网的迅猛发展, 数字图像的鉴别技术成为了研究的热点。图像的鉴别技术分为3类: 数字签名、数字水印和数字图像盲取证^[1]。其中, 数字签名和数字水印需要进行预处理, 是“主动”的取证技术, 数字图像盲取证不需要通过任何先前操作就可以检测图像是否经过篡改, 是一种“被动”的取证技术。由于现实性的要求, 盲取证技术越来越受到关注。

复制-粘贴是经常出现的一种篡改方式, 此方法就是将图像中的某块区域复制到其他区域以达到掩盖目的。王浩明等^[2]提出了块操作的思想, 设计了一种基于图像块的DCT系数的算法来检测图像复制伪造区域。赵俊红等^[3]提出了基于主成分分析的篡改方法。

上述具有代表性的算法在检测同幅图像复制-粘贴时取得了一定的效果, 但是在检测率和错误率上还不是理想, 有些算法对外界的抵抗性还不是很强。由于算法的检测率和错误率主要取决于选择的特征值, 笔者提出了一种基于二进小波

的检测算法来提高检测率、降低错误率, 并不采用不进行重叠分块的方式来提高整体检测时间效率。

1 二进小波

1.1 二进小波理论

一般来说, 在分析中使用傅里叶变换的地方均可以使用小波变换来进行代替, 但是小波分析在时域和频域里都具有比较好的局部化特性, 在分析对象的细节上更具有优势。

连续小波变换 (continue wavelet transform, CWT) 定义如下:

$$Wf(a, b) = \int_R f(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dt = \langle f, \psi_{a,b} \rangle. \quad (1)$$

对上式中的尺度参数 a 和平移参数 b 进行离散化处理, 一般选用 $a = 2^j, b = 2^j k$, 其中 $j, k \in \mathbf{Z}$, 变换后对于任意函数 $f(t) \in L^2(\mathbf{R})$ 对应的离散小波变换 (discrete wavelet transform, DWT) 的表达式为

$$Wf(j, k) = \int_R f(t) \frac{1}{\sqrt{2^j}} \psi\left(\frac{t}{2^j} - k\right) dt = \langle f, \psi_{j,k} \rangle. \quad (2)$$

在离散小波变换算法的快速实现过程中, 当进

收稿日期: 2014-05-08

基金项目: 国家自然科学基金项目 (61262075); 广西研究生教育创新计划项目 (YCSZ2013078)

作者简介: 张瑞芳 (1977—), 女, 硕士, 讲师, 研究方向: 图像处理, tdrf@sina.com.cn.

引文格式: 张瑞芳, 程晓辉, 宋子航, 等. 融合灰色马尔科夫理论的二进小波图像的复制-粘贴篡改检测算法 [J]. 桂林理工大学学报, 2014, 34 (4): 775-781.

行到采样的操作步骤时,会使得图像边缘的很多奇异点丢失,而当反过来进行图像重构的时候会出现 Gibbs 现象^[4]。视觉上看就是图像边缘的失真。在离散小波变换中,如果能做到变换中具有平移不变性的这个特性,就能解决上述问题,这就是二进小波变换。

在式(1)的小波变换中,令 $a = 2^j, j \in \mathbf{Z}, b$ 仍然跟连续小波变换中一样取连续变化的值,对于任意的函数 $f(t) \in L^2(\mathbf{R})$,二进小波变换可以定义为

$$Wf(2^j, b) = \int_{\mathbf{R}} f(t) \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-b}{2^j}\right) dt = \langle f(t), \psi_{2^j, b}(t) \rangle. \quad (3)$$

式中, $\psi_{2^j, b}(t) = \psi_{2^j}(t-b) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-b}{2^j}\right)$ 。

连续小波变换对尺度参数 a 和平移参数 b 都保持了连续性,相反离散小波变换对这两个参数都进行了离散化,而二进小波变换恰恰只是对尺度参数进行了离散化,在时间域上的平移参数仍保持了连续性,这也是它与离散小波变换的主要区别。

1.2 二进小波在本文算法中的应用

设 I 为待检测图像, $h[k]$ 和 $g[k]$ 分别为低通滤波器和高通滤波器,当 $j = 0$ 时,令 $I^0 = I$,当 $j = 1, 2, \dots, J$ 时,其相应的系数计算方法可以由下式得到:

$$c^{j+1}[n] = \sum_k h[k] c^j[n + 2^j k]; \quad (4)$$

$$d^{j+1}[n] = \sum_k g[k] c^j[n + 2^j k]. \quad (5)$$

而滤波器 $h^j[k]$ 和 $g^j[k]$ 分别由滤波器 $h[k]$ 和 $g[k]$ 在每个样本间插入 $2^j - 1$ 个 0 后得到。二进小波变换滤波过程为:当 $j = 0$ 时, $I^0 = I$,当需要 I^j 和 D^j 的尺度函数系数和小波函数系数的时候可以分别用 $h^{j-1}[k]$ 和 $g^{j-1}[k]$ 对 I^{j-1} 进行滤波得到。相应的二维二进小波分解如图 1 所示。

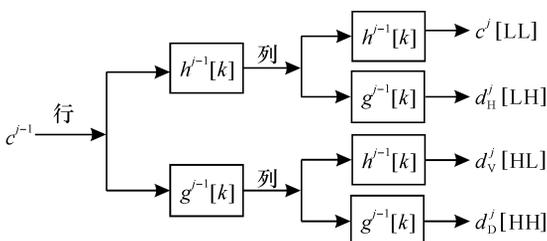


图 1 二维二进小波一级分解

Fig. 1 Two-dimension dyadic wavelet first level factoring

当图像进行完二进小波一级分解后将得到 4 个与原图大小一样的子图: LL、LH、HL、HH。其中 LL 子图为原图的逼近图像, HH 子图中包含了大量高频信息,算法将选取这 2 个子图进行分析。

2 检测算法

2.1 图像分块

对图像进行二进小波一级分解后,选取 LL 和 HH 子图进行图像分块。由于一般篡改区域不会很小,一次移动一个像素没有任何必要。在本文的算法中,将不进行滑动移动来进行图像块的划分,直接将图像划分成 $R \times R$ 彼此相邻的图像块(或者说步长为 R),同样对于 $M \times N$ 的图像,经过这样的划分后,图像块的个数为 $\lceil M/R \rceil \times \lceil N/R \rceil$ 。在这里,选择图像块的大小为 16×16 来进行整个图像的分块,即保证某些图像块完全包含了图像篡改区域的一部分。

2.2 相似区域匹配

设 LL 和 HH 子图中每个图像块的大小为 $R \times R$,全部转换成 $1 \times R^2$ 的矩阵形式(即每一图像块转换成一行),每个图像块作为一个向量,LL 和 HH 子图中的所有图像块转换成一个矩阵。

定义向量间距离公式为

$$d(p, q) = \sqrt{\frac{1}{N} \sum_{i=1}^n (p_i - q_i)^2}. \quad (6)$$

式中, N 取 256,也就是将向量之间的距离进行规范化,相应的距离数值转换到 $0 \sim 1$,以便进行阈值的设定。在进行字典排序之前,定义 2 个阈值 Th_1 和 Th_2 。由于 LL 子图为原图的逼近图像,通过字典排序算法,保存 LL 子图中所有距离 $d < Th_1$ 的图像块在表 Llist 中。而 HH 子图中主要为图像块的噪声和边缘等高频信息,一旦图片经过篡改,这些敏感信息必定会发生改变,保留 HH 子图中所有距离 $d > Th_2$ 的图像块在 Hlist。最终在保留下来的图像块对中,挑选出图像相似块对。从 K_{LL} 和 K_{HH} 树中分别选取出来的图像块对列为 2 个表 Llist 和 Hlist,如果图像块对同时出现在 2 个表中,此图像块即被判定为相似图像块。

2.3 篡改区域确定

通过相似区域匹配找到相似图像块后,由于并不是进行重叠的图像分块方法,有些图像块中

可能只包含一部分篡改区域而无法达到相似度要求从而被检测出来,而检测出来的图像块基本上都是全部包含了篡改区域的,故需要进一步确定篡改区域。下面介绍一种图像块的边缘处理方法^[5]。

图2中一个方格代表图像中的一个图像块。假设 X 和 Y 部分所组成的图像块为前面步骤得到的相似图像块,即 X 边缘所包含的区域。具体步骤如下:

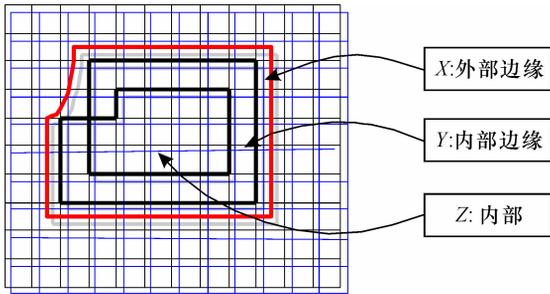


图2 篡改区域图像边缘处理

Fig. 2 Tampering with regional image edge processing

(1) 确定图像块的内部边缘。对于先前所得到的相似图像块,对每个检测出来的相似图像块进行4邻域(即上下左右)图像块的判断,如果4邻域也都还在篡改图像块里,则把它划分为 Z 区域(内部),如果4个邻域中存在1~3个区域为图像块,就把它划分为 Y 区域(内部边缘),4邻域不存在的图像块不作任何处理。

(2) 外部边缘的确定。把 Y 区域(内部边缘)中的图像块的所有8邻域(即组成包围该图像块的所有图像块)图像块标出,已经存在于 Y 和 Z 区域的(即先前已检测出的篡改区域图像块)不作任何处理,剩下的全部作为 X 区域(外部边缘)。

(3) 篡改区域的完善。对于前面计算得到的内部边缘中的图像块,按照步长为1的像素移动来进行图像块的重叠划分,位移的方向由 Y 指向 X ,并且都在 X 之内。通过移动划分得到的图像块按照2.2节中相似区域确定的方法来进行判断,最终确定篡改区域。

2.4 算法流程

(1) 对于待测图像 I ,设其大小为 $M \times N$,首先将其转换到256级的灰度空间;

(2) 对图像进行二进小波一级分解,变换后提取出LL和HH子图;

(3) 对LL和HH子图分别进行无重叠图像分块,每个图像块的大小为 16×16 ,每个图像块相邻;

(4) 对于LL和HH子图,每个子图里的每个图像块的特征值排成一行,一起构建成特征向量矩阵,然后构建LL和HH的K-D树 K_{LL} 和 K_{HH} ;

(5) 设定阈值 Th_1 和 Th_2 。对于 K_{LL} 树,通过搜索每个结点距离小于 Th_1 的图像块对,保存在Llist表中;而对于 K_{HH} 树,通过搜索每个结点距离大于 Th_2 的图像块对,保存在Hlist表中;

(6) 选取Llist和Hlist表中都存在的图像块对,作为篡改图像块;

(7) 进行图像块边缘处理,进一步完善篡改区域;

(8) 算法结束。

3 灰色-马尔科夫滤波修正模型

在进行实验分析前,由于要对上百个数据进行比较,因而需对数据进行预测。目前常见的预测理论有GM(1,1)预测、马尔科夫链、回归分析预测和德尔菲法等^[6]:德尔菲法是建立在专家选定的基础上的主观方案,具有很强的主观性^[7];灰色模型主要适用于单一指数增长的序列,对于异常波动的序列通常是无能为力;回归分析则考虑多种因素,增加了计算工作量,建模相对较困难^[8]。

针对上述问题,需要在灰色预测的基础上,结合马尔科夫理论,建立融合马尔科夫修正因子下的灰色预测模型,从而能够有效地将灰色理论的弱化序列波动性优点和马尔科夫模型状态预测特性等优点结合。

3.1 预测模型GM(1,1)

假定待预测数据集中的原始数据为

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n));$$

对 $x^{(0)}$ 作1次累加得到新的数列

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)), \quad (7)$$

其中 $x^{(1)}(k) = \sum_{i=0}^{k-1} x^{(0)}(i), k = 1, 2, \dots, n$,求均值数列 $z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), k = 2, 3, \dots, n$,建立GM(1,1)的白化方程为

$$dx^{(1)}/dt + ax^{(1)}(t) = b, \quad (8)$$

$$Y_1 = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, \quad B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix},$$

则由最小二乘法,求使得 $J(u) = (Y_1 - B \cdot \hat{u})^T (Y_1 - B \cdot \hat{u})$ 达到最小值的 \hat{u} 。求解方程(8)得到预测值 $x^{(1)}(k+1) = (x^{(0)}(1) - b/a)e^{-ak} + b/a$,

$$k = 1, 2, \dots, n-1. \quad (9)$$

式(9)是模型的 $x^{(1)}(k+1)$ 的预测值,对 $x^{(1)}(k+1)$ 进行邻近数据累减,得到 $x^{(0)}(k+1)$,即

$$\begin{aligned} x^{(0)}(k+1) &= x^{(1)}(k+1) - x^{(1)}(k) \\ &= (x^{(0)}(1) - a/b)(e^{-a} - 1)e^{-a(k-1)}. \end{aligned} \quad (10)$$

3.2 马尔科夫预测模型

马尔科夫链具有无后效性^[9],链中节点某一时刻所处的状态只与该状态前的状态有关,而与其后的状态无关。马尔科夫的预测理论主要是以现在所处节点,预测未来等距节点发生的概率。

(1) 状态划分。在马尔科夫预测模型中,状态划分是一个很重要的过程。依照系统目前状态来预测各个目标未来可能发生的概率,因而组成一个有限状态集^[10]。

(2) 状态转移概率。在事件的发生过程中,从状态 S_i 出发,进而转移到状态 S_j 的可能性,称为状态转移概率^[8],记为 P_{ij} 。在转移过程中, S_i 出现的次数为 n_i ,状态 S_i 转到状态 S_j 次数为 n_{ij} ,则 $P_{ij} = n_{ij}/n_i$ 。

(3) 状态转移概率矩阵。某一预测事件中,在 S_1, S_2, \dots, S_n 里共有 n 个可能状态^[11],那么从状态 S_i 到 S_j 转移概率,将所有的可能概率组成一个概率矩阵

$$P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix}, \text{ 并且 } P_{ij} \text{ 满足条件}$$

$$\begin{cases} 0 \leq P_{ij} \leq 1, i, j = 1, 2, \dots, n; \\ \sum_{j=1}^n P_{ij} = 1, i = 1, 2, \dots, n. \end{cases}$$

(4) 预测值。在上几步中,确定未来的状态转移概率后,预测值变动区间 $[\mu_0, \mu_1]$ 也就确定,使用变动区间的中点^[12],可得到 $h(t) = (\mu_0 + \mu_1)/2$ 。

3.3 灰色-马尔科夫模型

灰色-马尔科夫模型融合灰色系统和马尔科夫理论,模型主要分为2步:①在灰色系统理论上增加了对背景数据的光滑处理^[13]以及数据的扩充,进而降低数据本身对建模的影响,同时

$z^{(1)}(k)$ 也进行了调整,并不再是单一均值,而是考虑到前后数据内在关联;②第①步生成的预测值作为 GM(1, 1) 的输入值,进行马尔科夫预测。假设已测得数据项为 P 个,预测项 m 个。

(1) GM 滤斗预测。与传统 GM 模型不同,GM 滤斗模型的原始数列为 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$,并且运用对数对 $x^{(0)}(k)$ 进行滤斗处理,即 $x^{(0)}(k) = \ln X(k)$, $X(k)$ 为原始数据。滤斗后得到的最后预测值 $\hat{x}^{(0)}(k+1) = (x^{(0)}(1) - a/b)(e^{-a} - 1)e^{-a(k-1)}$,可通过 $\ln(\hat{x}^{(0)}(k+1)) = \hat{x}^{(0)}(k+1)$ 求出改进的 GM 预测值 $\hat{x}(k+1)$ 。

(2) 马尔科夫修正。在马尔科夫修正的过程中,模型对 GM 滤斗模型预测的数据进行再处理,生成马尔科夫修正序列 $m(k)$, $m(k) = 0.5\hat{x}(2k-1) + 0.5\hat{x}(2k)$, $k = 1, 2, \dots, 9$,并将 $m(k)$ 划分成为多个状态,再根据预测值 $m(k)$ 所处状态 j 以及前一个状态 i ,计算预测的马尔科夫状态的修正值

$$\mu(k) = P_{ij}h(j). \quad (11)$$

最终第 i 预测值为 $f_i(k) = m(k) + \mu(k)$ 。

(3) 模型计算步骤:①用 GM(1, 1) 滤斗模型根据前 p 个已知数据项,计算 $p+i$ 个预测结果;②根据①中的 $p+i$ 的预测结果计算预测误差值,将预测的误差按规则划分马尔科夫转移状态,并计算转移概率矩阵;③利用公式 $f(k) = m(k) + \mu(k)$ 计算第 $p+i$ 个数据项;④将第 $(i+1) \sim (p+i)$ 个数据序列作为输入项,使用 GM(1, 1) 滤斗模型预测 $p+i+1$ 个待测项,并将得到的结果作为输入返回到②,直到所有待测项全部预测完毕。模型流程图图3。

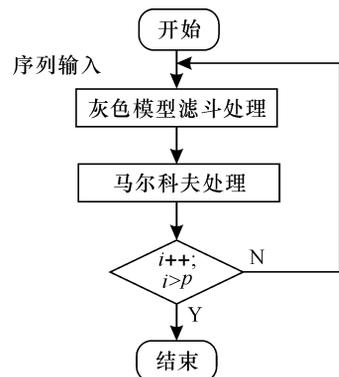


图3 模型流程

Fig. 3 Process model

4 GM - 马尔科夫预测及实验结果

在实验过程中，定义检测率 r 与错误率 w 为

$$r = \frac{\sum_i |A_i \cap P_i| + |B_i \cap Q_i|}{\sum_i |A_i| + |B_i|}, \quad (12)$$

$$w = \frac{\sum_i |A_i \cup P_i| + |B_i \cup Q_i|}{\sum_i |A_i| + |B_i|} - r. \quad (13)$$

式中： A_i 和 B_i 为待测图像中的复制和粘贴的区域，而 P_i 和 Q_i 则是被检测出来的区域。

在实验中使用灰度图像，如果为彩色图像则先转化成灰度图像再进行处理，实验图像大小为 320×320 ，原始图像通过使用 Adobe Photoshop CS 进行复制-粘贴篡改。对于参数 R ，经过大量的实验，当 R 选择较小数值时（比较常用的数值是 4、8），由于小图像块在某一些区域从统计特性来判断相似度很高，会产生很多误判，而常理上一般进行篡改，其篡改区域的大小往往比 8×8 区域大。当 R 选择较大数值的时候（比如选择 32、64），往往会使得检测率较低。在此，笔者选择的图像块模板大小为 16×16 ，也就是 R 值为 16。表 1 列出了分别使用不同大小的图像块进行算法检测时的检测率与错误率的对比。

表 1 图像块选择比对

Table 1 Comparison of detection rate and error rate among images of different sizes

图像大小	检测率/%	错误率/%
8×8	94.0	13.6
16×16	93.1	5.15
32×32	86.3	4.78

参数 Th_1 和 Th_2 的选择对于算法准确率有非常大的影响： Th_1 主要控制图像块的相似性匹配， Th_1 值的选择直接影响到算法检测率的高低，而在 HH 子图中反映的是图像的噪声和边缘等高频信息，由于同幅图像内一旦存在复制-粘贴操作将破坏噪声和边缘这些高频信息， Th_2 值选择的恰当可以滤去图像自身的未篡改的相似区域，即可以大大降低误判率。经过反复实验， Th_1 和 Th_2 的配合选择和误判率的关系如图 4 所示。

可以看出，随着阈值 Th_1 和 Th_2 的不断增大，误判率也迅速增大。若 Th_1 和 Th_2 选择的数值太小，虽然误判率小了，但实验表明已经检测不出篡改区域，即篡改检测率不高。通过误判率和检测率的综合比较，本算法最终选择 Th_1 的值为 0.71， Th_2 的值为 0.29，这样的组合将给检测带来最佳的检测率和误判率数值。

图 5 为一幅原始图像，图 6 为使用图像编辑软件对其进行了复制-粘贴型篡改，图 7 为使用本文

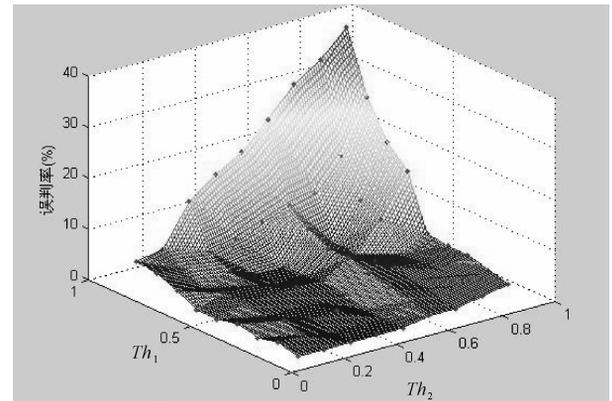


图 4 阈值 Th_1 和 Th_2 的选择与误判率的关系

Fig. 4 Selection and misjudgment rate relationship of Th_1 and Th_2



图 5 原始图像

Fig. 5 Original image



图 6 篡改图像

Fig. 6 Tamper image



图 7 检测结果

Fig. 7 Test results in figure

算法对其进行检测的结果图。由图7可知,存在2块白色区域,也就是所标出的复制、粘贴区域。

采用本算法选取10幅320×320的图进行篡改,预测其后490幅的检测率和错误率,最后进行平均检测率和错误率的计算,得到最终的平均检测率和平均错误率,这10幅图的统计数据见表2。

表2 检测结果
Table 2 Detection results

图序号 <i>k</i>	检测率/%	错误率/%
1	93.0	5.22
2	92.27	5.71
3	93.68	5.85
4	92.91	6.12
5	91.8	6.43
6	94.56	4.65
7	93.27	5.23
8	93.37	4.54
9	93.79	5.23
10	93.78	4.78

对数据作滤斗处理:数据随机性会导致曲线的抖动,故对原始测试数据取对数调整,降低数据抖动对模型的影响。建立数据列

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)),$$

调整后数据进行1次累加计算,则 $x^{(1)}(1) = x^{(0)}(0)$,

$$x^{(1)}(i) = \sum_{k=1}^i x^{(0)}(k) \quad (i = 2, 3, \dots, 10),$$

由最小二乘法求得 $\hat{u}^r = (B^T B)^{-1} B^T Y_n = \begin{bmatrix} -0.0017 \\ 4.9936 \end{bmatrix}$ (检测率),

$$\hat{u}^w = (B^T B)^{-1} B^T Y_n = \begin{bmatrix} -0.0274 \\ 1.2402 \end{bmatrix}$$
 (错误率)。

最后根据上述理论分析所建立的滤斗后的GM(1,1)模型,可以分别得到检测率、错误率的预测值

$$\begin{cases} \hat{x}^{(1,r)}(k+1) = 2938.342e^{0.0017k} - 2937.412 \\ \hat{x}^{(0,r)}(k+1) = \hat{x}^{(1,r)}(k+1) - \hat{x}^{(1,r)}(k) \end{cases},$$

$$\begin{cases} \hat{x}^{(1,w)}(k+1) = 45.31e^{0.0274k} - 45.26 \\ \hat{x}^{(0,w)}(k+1) = \hat{x}^{(1,w)}(k+1) - \hat{x}^{(1,w)}(k) \end{cases},$$

$k = 1, 2, \dots, 10$ 。检测率与错误率的预测值与实际值如表3、表4所示。

运用马尔科夫链分析方法分别对改进后的GM预测结果进行分析,可将状态划分为6种,即

A: 实际值与预测值差在 $[-3, -2]$;

B: 实际值与预测值差在 $[-2, -1]$;

表3 检测率 GM 预测结果

Table 3 Predicted results of detection rate GM

图序号 <i>k</i>	实际值/%	预测值/%	误差	状态
1	93.0	93.78	0.78	D
2	92.27	93.13	0.86	D
3	93.68	94.63	0.95	D
4	92.91	94.18	-1.27	B
5	91.8	94.39	-2.59	A
6	94.56	94.87	-0.31	C
7	93.27	93.46	-0.19	C
8	93.37	92.01	1.36	E
9	93.79	93.54	0.25	D
10	93.78	96.53	-2.75	A

表4 错误率 GM 预测结果

Table 4 Predicted results of error rate GM

图序号 <i>k</i>	实际值/%	预测值/%	误差	状态
1	5.22	4.36	0.86	D
2	5.71	4.51	1.2	E
3	5.85	4.89	0.96	D
4	6.12	7.42	-1.3	B
5	6.43	8.67	-2.24	A
6	4.65	4.27	0.38	C
7	5.23	5.13	0.1	C
8	4.54	6.12	-1.58	D
9	5.23	4.45	0.78	D
10	4.78	7.64	-2.86	A

C: 实际值与预测值差在 $[-1, -0]$;

D: 实际值与预测值差在 $[0, 1]$;

E: 实际值与预测值差在 $[1, 2]$;

根据以上分类,分别得到状态转移矩阵

$$S^r = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, S^w = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

通过转移矩阵可以得出马尔科夫转移概率矩阵

$$P^r = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

$$P^w = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0.25 & 0.25 & 0 & 0.25 & 0.25 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

根据马尔科夫预测模型原理可知最后2组预测值,性能分析见表5。可知GM-Markov模型的预

表5 检测率预测模型误差

Table 5 Error in predicted model of detection rate

图序号 k	实际值 /%	GM 预测模型		GM-Markov	
		预测值/%	误差	预测值/%	误差
9	93.79	93.54	0.25	93.69	0.1
10	93.78	96.53	-2.75	94	-0.22

测精度很高,为计算方便,针对100组数据进行取均值,先以目前10组数据的变化模型为依据,进行检测率与错误率的90组数据的预测,其结果取其叠加之后的平均值,经测试与文献[2]中算法比较如表6。

可以看出,本文算法在大量检测数据支持的情况下,平均检测率有了提升,且平均错误率有了一定下降,这主要取决于在图像进行二进小波分解后对LL和HH子图的综合分析。

表6 检测结果比较

Table 6 Test results comparison

	本文算法	文献[2]算法	%
平均检测率	93.68	92.95	
平均错误率	6.32	7.05	

5 结论

本文提出的基于二进小波图像复制-粘贴篡改的检测算法,通过进行不重叠的图像分块,对图像进行二进小波分解后选取低频和高频子图来分析。利用滤斗式灰色-马尔科夫模型取得多组数据的预测,实验结果的可靠性、真实得以提性高。结果证明本文算法可提高检测率并降低了错

误率。但是本算法还存在一定不足,不能抵抗旋转、缩放,对于有损、压缩与已处理图像再滤波等类型的检测,尚不能处理,这也是下一步研究的方向。

参考文献:

- [1] 曹刚. 数字图像操作取证技术研究 [D]. 北京: 北京交通大学, 2013.
- [2] 王浩明, 杨晓元. 一种基于 DCT 系数直方图差异的 JPEG 图像篡改检测 [J]. 四川大学学报: 工程科学版, 2014, 46 (1): 41-46.
- [3] 赵俊红, 康文雄. 基于投影数据主成分分析的图像篡改检测算法 [J]. 计算机工程, 2012, 38 (10): 203-205.
- [4] 姜闪闪. 基于二进小波变换的自适应阈值图像去噪研究 [D]. 乌鲁木齐: 新疆师范大学, 2010.
- [5] 张广辉, 汪杭军. 基于 FMT 的快速 Copy-Move 篡改检测 [J]. 计算机工程与设计, 2010, 31 (15): 3530-3532.
- [6] 侯丽敏, 马国峰. 基于灰色线性回归组合模型铁路客运量预测 [J]. 计算机仿真, 2011 (7): 9-11.
- [7] 尹琦, 袁小平. 基于灰色预测可变裕度 PID 网络自适应算法 [J]. 计算机应用研究, 2012 (2): 317-319.
- [8] 于波, 于东, 孙建伟. 马尔科夫决策过程在多路径冗余传输调度算法中的应用 [J]. 小型微型计算机系统, 2012 (4): 177-181.
- [9] 赵玲, 许宏科. 基于等维新息灰色马尔科夫模型的互联网用户人数预测研究 [J]. 计算机科学, 2013 (4): 125-127.
- [10] 崔兴凯, 路秀英. 基于 NGM(1,1,k) 模型的农产品产量预测方法 [J]. 微电子学与计算机, 2011 (8): 207-209.
- [11] 蔡明亮. 改进的灰色马尔科夫模型及其对全国邮电业务总量的预测 [D]. 南京: 南京邮电大学, 2013.
- [12] 高蔚. 基于 Markov 理论的改进灰色 GM(1,1) 预测模型研究 [J]. 计算机工程与科学, 2011 (2): 163-167.
- [13] 李小力, 李言俊, 张科. 改进的灰色预测模型在导弹中的应用 [J]. 计算机仿真, 2010 (8): 39-42.

Grey-Markov theory in blind detection of copy-move forgery in digital images by dyadic wavelet transform

ZHANG Rui-fang¹, CHENG Xiao-hui¹, SONG Zi-hang¹, CHEN Yang²

(1. College of Information Science and Engineering, Guilin University of Technology, Guilin 541004, China; 2. Guilin Municipality Public Security Bureau, Guilin 541001, China)

Abstract: A detection algorithm is put forward in dyadic wavelet image which made binary wavelet image decomposition with original image. The feature of each image block is extracted in high and low frequency and analyzed using edge processing. For the correctness during the tamper detection in a large number of images, a type of fusion model is put forward based on the grey theory and Markov chain to predict the binary image detection rate and error rate of copy-paste tampering algorithm, to obtain the average success rate of the algorithm by computing the average value of the predicted value. Comparison of literature about the mean value of detection and error rate proves that in the condition of avoiding accidental circumstances, the algorithm can improve the detection rate and reduce the error rate.

Key words: dyadic wavelet; copy-move forgery; edge processing; grey Markov chain