

# 基于人工神经网络的数字图像盲取证

高 强, 张华熊

(浙江理工大学信息电子学院, 杭州 310018)

**摘 要:** 数字图像盲取证技术是一项通过图像本身统计特性来鉴别图像是否被篡改的技术。文章基于数码相机的成像原理,利用人工神经网络来预测成像过程中由于 CFA(color filter arrays)插值算法导致图像像素点之间出现的相关性关系,由于没有篡改的图像采用的 CFA 插值算法是相同的,而被篡改的图像则可能存在两种以上的 CFA 插值算法,两种情况下图像预测误差会出现差异,本算法利用这种差异实现数字图像的盲取证。实验结果表明本算法具有一定的有效性。

**关键词:** 数字图像盲取证; CFA 滤波阵列; 人工神经网络

**中图分类号:** TP368.2      **文献标识码:** A

## 0 引 言

随着数码照片在新闻、司法、文艺等各个行业的大量使用,鉴定数码照片的真实性已成为采用照片前的重要工作。然而,由于数码照片与生俱来的数字特性使其很容易被 Photoshop 等软件修改且很难被觉察。目前,比较成熟的图像防伪方法是嵌入水印,一旦图像被篡改,图像嵌入的水印也会随之变化,以此可鉴定图像的真伪。然而这种方法是牺牲图像视觉效果为代价的,图像在被加入水印的同时也被破坏了原有的视觉效果。

鉴于水印技术的缺陷,数字图像盲取证技术吸引了越来越多研究者的兴趣。在数字图像盲取证的过程中图像数据不需要嵌入其他信息,仅通过图像自身的统计特性来鉴定图像是不是被篡改过。因此,该方法不仅达到了鉴定图像真伪的目的,也不需要再在图像中加入额外信息,保证了图像原有的视觉效果。数字图像盲取证的方法有很多种,有的利用不同相机拥有不同噪声的特点,通过对图像噪声的统计信息来鉴别图像是否被篡改<sup>[1]</sup>;有的利用在图像特征点周围提取特征向量,通过特征向量的欧氏距离来鉴别图像是否被篡改<sup>[2]</sup>;有的则根据不同相机使用不同的 CFA 颜色插值算法来鉴别图像的真伪<sup>[3-6]</sup>,本文所讨论的数字图像盲取证算法就属于这种鉴定方法的范畴。

## 1 CFA 颜色滤波阵列原理

### 1.1 CFA 颜色滤波阵列

现代数码相机一般采用电荷耦合器件(CCD)作为感光元件,在 CCD 芯片表面整齐地排列了一个由微小感光单元构成的阵列,当光线中的光子撞击阵列上的感光单元时,感光单元产生电子,每个感光单元所产生的电子数目又与撞击每个感光单元的光子数目成正比,这样,当一幅照片曝光之后光子数目就可以从 CCD 感光阵列中读出来,从而形成一幅影像。然而,这样的一幅影像并不包含真实图像的颜色信息,是一幅单色图,为了形成彩色照片还需要在 CCD 元件前安装一块由红、绿、蓝三种颜色交替出现所形成的滤波阵列,自

然光通过颜色滤波阵列后在 CCD 感光单元上投射出红、绿、蓝三种颜色交替的光子,这样从每个感光单元所产生的信息中只能提取出一种颜色的图像值,而另两种颜色信息则被丢弃。为了获得缺失的颜色信息还需通过特定的插值算法根据周围其他具有该颜色的感光单元估算出来。经过对每个感光单元数据的估算最终形成一幅完整的彩色图像<sup>[7]</sup>,如图 1 所示。

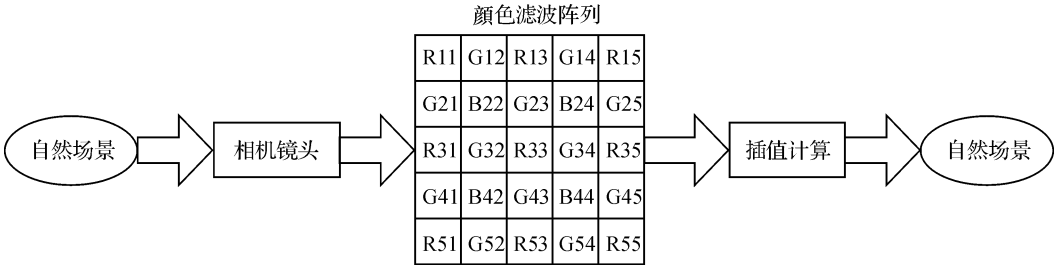


图 1 单 CCD 芯片数码相机成像原理

1.2 CFA 插值算法

插值算法是一种根据周围像素点数据估算自身像素数据的一种算法,在图像放大,图像拼接及上文提到的数码照片生成中有广泛的应用,根据估算数据与被估算数据像素点的位置关系及运算算法大致可以分为:双线性、平滑色调、中值滤波、最近临点等插值算法。

插值图像的估算过程如下。

a)  $S(x,y)$ 代表反射进入照相机镜头的原图像,该图像经过 CFA 滤波阵列后,经过红、绿、蓝三种颜色滤波单元的像素点分别被滤波为  $\tilde{R}(x,y)$ 、 $\tilde{G}(x,y)$ 、 $\tilde{B}(x,y)$  三幅单色图中的一个像素,而原图像每个像素的另外两种颜色值被丢弃,置为 0。如公式(1~3)所示。

$$\tilde{R}(x,y)=\begin{cases} S(x,y), & \text{如果 } S(x,y)=r_{x,y} \\ 0, & \text{其他点} \end{cases} \tag{1}$$

$$\tilde{G}(x,y)=\begin{cases} S(x,y), & \text{如果 } S(x,y)=g_{x,y} \\ 0, & \text{其他点} \end{cases} \tag{2}$$

$$\tilde{B}(x,y)=\begin{cases} S(x,y), & \text{如果 } S(x,y)=b_{x,y} \\ 0, & \text{其他点} \end{cases} \tag{3}$$

这时三幅滤波后的图像叠加后就成为一幅每个像素点的  $R$ 、 $G$ 、 $B$  值只有一个为原始图像的实际值而其他两种颜色值为 0 的图像。

b) 依照设定的插值算法,根据每个像素点所缺失的颜色,利用该点周围具有该缺失颜色的像素点的颜色值估算出该像素点缺失的颜色。 $h_r(u,v)$ 、 $h_g(u,v)$ 、 $h_b(u,v)$ 代表像素点插值算法的运算系数,利用该系数与周围多个含有本像素点缺失颜色的像素点颜色值相乘并求和得到该像素的缺失颜色估算值  $R(x,y)$ 、 $G(x,y)$ 、 $B(x,y)$ 。如公式(4~6)所示。

$$R(x,y)=\sum_{u,v=-N}^N h_r(u,v)\tilde{R}(x-u,y-v) \tag{4}$$

$$G(x,y)=\sum_{u,v=-N}^N h_g(u,v)\tilde{G}(x-u,y-v) \tag{5}$$

$$B(x,y)=\sum_{u,v=-N}^N h_b(u,v)\tilde{B}(x-u,y-v) \tag{6}$$

1.2.1 最近临点插值算法

最近临点插值算法认为图像中相邻像素点之间的颜色值比较接近,因此当一个像素点缺少某一颜色通道的颜色值时可以取最近临点中具有该颜色的颜色值近似的作为该点的缺失颜色值。

1.2.2 双线性插值算法

在双线性插值算法中每个像素点的缺失颜色是以该像素点  $3\times 3$  的邻域内具有该颜色的像素点的颜色值的平均值确定的,因此确定邻域中含有该缺失颜色的像素点位置后,将这些具有缺失颜色的点的颜色数据

取平均即得到该像素点的颜色值。

### 1.2.3 平滑色调插值算法

平滑色调插值算法认为在一定范围内, R、G、B 三种颜色在图像中的变化比例相近, 则对于两个相邻的点  $(X, Y)$ 、 $(M, N)$ , 有  $R(X, Y)/R(M, N) = G(X, Y)/G(M, N) = B(X, Y)/B(M, N)$ 。由于定义在某一范围内红、蓝两色与绿色的比值不变, 要求  $R(X, Y)$  的值, 则有  $R(X, Y) = G(M, N)/G(X, Y) * R(M, N)$ 。

因此在平滑色调插值算法中, 首先通过双线性差值算法算出整幅图的 G 通道的颜色值, 然后根据上述方法得到图像中其他两种颜色的颜色值。

### 1.2.4 中值滤波插值算法

中值滤波插值算法首先通过双线性差值算法计算出每个像素点的 R、G、B 三个颜色通道的颜色值, 然后对 R、G、B 三个通道的颜色值进行两两相减, 再对获得的差值进行中值滤波, 最后通过滤波后的差值与各点实际获得的颜色值相加得到。

若  $C_{rg}(x, y) = R(x, y) - G(x, y)$ ,  $C_{rb}(x, y) = R(x, y) - B(x, y)$ ,  $C_{gb}(x, y) = G(x, y) - B(x, y)$  分别表示 R、G、B 三个通道颜色值的两两差值, 则  $M_{rg}(x, y)$ 、 $M_{rb}(x, y)$ 、 $M_{gb}(x, y)$  分别表示  $C_{rg}(x, y)$ 、 $C_{rb}(x, y)$ 、 $C_{gb}(x, y)$  滤波后的值。

经过上述几种插值算法的插值计算找回了图像中原本缺失的颜色信息, 再通过 R、G、B 3 个颜色通道的数据叠加后就得到一幅完整的图像。

## 2 BP 神经网络图像盲取证算法

由于相机对图像缺失颜色的复原是通过插值算法实现的, 插值算法的优劣决定了相机对真实色彩的还原效果, 因此各个数码相机生产厂家都在不断地改进自己的颜色插值算法, 最终形成了不同数码相机具有不同插值算法的结果。

基于这一情况, 本文提出了基于人工神经网络的图像盲取证算法, 即通过人工神经网络逼近图像的插值算法并估算出像素点的缺失颜色值, 然后与待鉴定图像各像素点的实际颜色值进行比对。由于不同相机使用的是不同的插值算法, 因此当一部数码相机所拍摄的图像与另外一部相机拍摄的图像拼接篡改之后, 该幅图像就包含了两种插值算法所估算出来的图像数据, 假设造假者只篡改了原图的一小部分, 则在该篡改部分中, 通过原图估算出来的颜色值将会与伪造图像的实际值有较大差异, 考虑到相机噪声和图像不同存储格式的压缩算法所引入的噪声, 取固定面积内图像实际颜色值与估算值误差的平均值, 超出平均值的图像部分就认为是被篡改过的。大部分相机所使用的插值算法都是非线性的, 因此选用非线性逼近效果比较好的 BP 神经网络来进行计算。

算法的具体运算步骤为: a) 取出图像红色通道的数据, 将红色通道图像等分为如图 2 所示的  $3 \times 3$  共 9 份子图  $ti$ , 假设图像的篡改者只修改的图像中较小的一部分, 那么该篡改区域不会出现在所有的子图中, 因此依次利用 9 份子图中的一幅  $ti$ , 训练 BP 神经网络  $BPNET_{ti}$ , 然后利用该网络估算出 9 份子图所缺失的红色分量  $BPTi_i$ 。如公式(7)、(8)所示。

$$BPNET_i = BP(ti) \quad (7)$$

$$BPTi_i = BPNET_i(ti) \quad (8)$$

b) 求出 9 份子图的估计值与实际图像红色值的误差阵列  $DIFF_i$  并求出每份子图各像素点的误差平均值, 以训练 BP 网络所用子图的平均误差值为判决标准, 与其余子图误差平均值做差, 然后以  $\pm 35$  为判决门限判断其余子图是否与本子图误差接近, 每发现一个与本身误差平均值相近的子图就将本子图的权值加 1。依次对 9 幅子图进行同样运算最后得到权值最高的子图, 则认为该子图是未被篡改图像污染的。如公式(9)所示。

$$DIFF_i = ti - BPTi_i \quad (9)$$

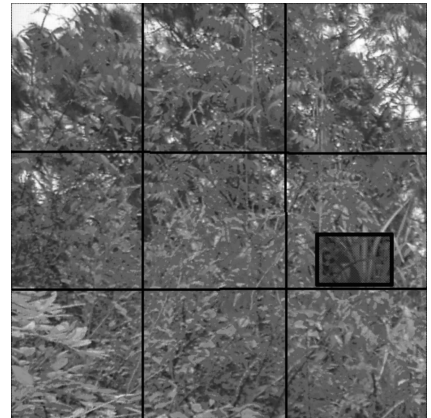


图2 原始图像

c)以该未被污染子图训练的 BP 神经网络估算整幅图像,得到整幅图像红色分量的估算值与实际值误差的空间分布图,在误差空间分布图中大误差点集中出现的区域就是被篡改的区域。

3 实验结果及分析

由于本算法使用训练 BP 网络所用子图的估算值与实际值的平均误差作为判决标准,因此 BP 网络能否对待检测图像的插值算法提供精确的估算决定了本算法对篡改图像的鉴定效果。本文对每种差值算法获得的 60 幅图像利用自身数据训练 BP 网络然后估算自身数据,并将误差值落在 $\pm 35$ 区间内的平均值逐个记录下来最后计算该均值的平均值。从表 1 中可以看出,随着插值算法复杂度的提高 BP 网络估算图像数据的能力随之下降。

表 1 60 幅图各种插值算法下自身估算值平均误差

| 插值算法  | 平均概率/% |
|-------|--------|
| 最近临点  | 98.4   |
| 双线性差值 | 97.8   |
| 色彩平滑  | 73.3   |
| 中值滤波  | 52.7   |

为了更清晰地反映算法效果,本文首先生成 4 幅分别使用最近临点、双线性、色彩平滑、中值滤波算法插值出来,包含同一景物信息的图像,如图 2 所示。然后依次利用其中一种插值算法训练神经网络后去估算这 4 幅图像,得到每种差值算法的误差概率密度分布图。以此说明通过 BP 神经网络的估算和概率密度的统计可以清晰地反应不同的插值算法所插值出来的图像的差别。

对篡改图像进行鉴定。分别将这 4 幅图像在图 2 黑框内的部分篡改为其他 3 幅图像中该部分的图像,并将伪造图像送入本文的算法程序进行计算,得到被篡改图像的估算值与实际图像值之间的误差空间分布图。

从实验数据图 3~图 6 中不难看出,对于由最近临点、双线性、中值滤波 3 种插值算法得到的图像,使用与自身相同插值算法估算出来的像素点颜色值与实际值之间的误差较小,且这些误差较小的点落在误差区间内概率密度较大,而使用与自身不同插值算法估算出来的图像与实际值之间的误差较大,且这些误差点的概率分布较为分散。反映到鉴定被篡改的图像中,图像被篡改过的区域的估算值与实际图像值之间的误差明显大于未被篡改部分的误差,在单位面积内误差平均值要明显高于未被篡改部分,如图 3(e)、图 4(e)、图 5(e)所示。

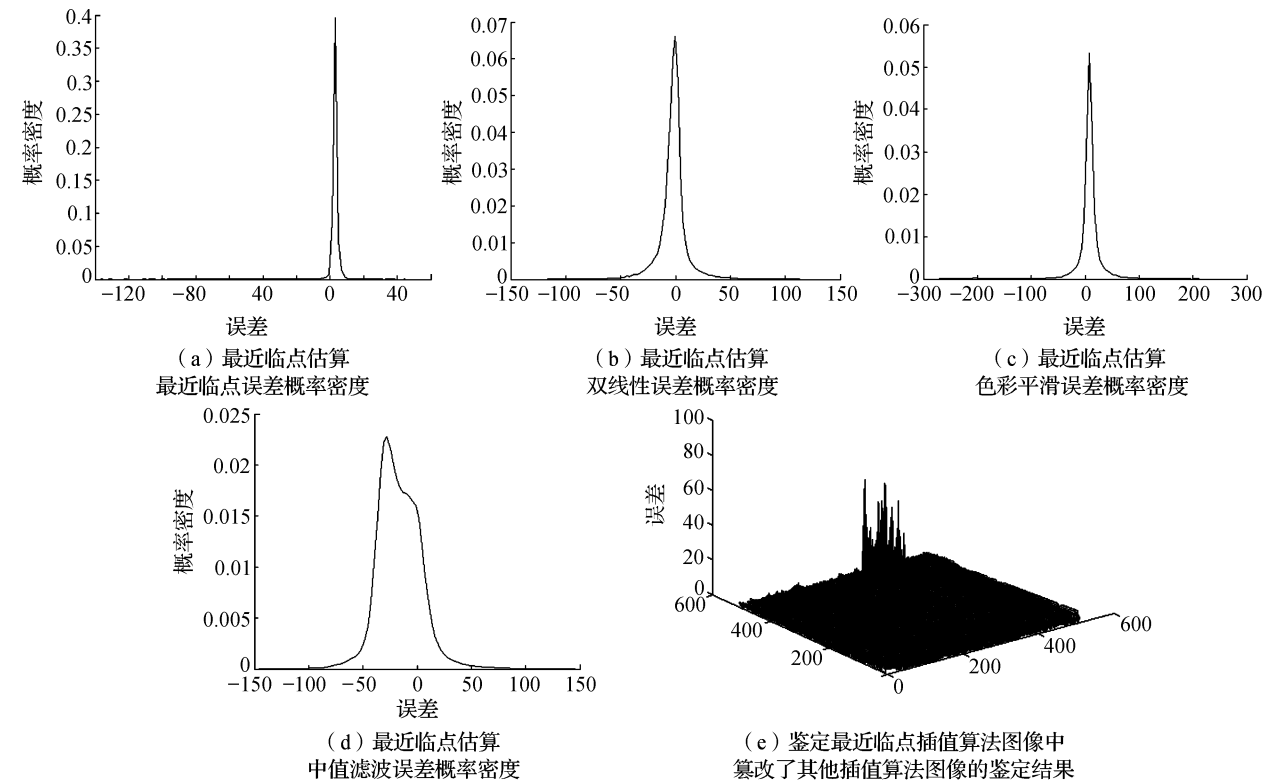


图 3 对最近临点插值算法图像试验结果

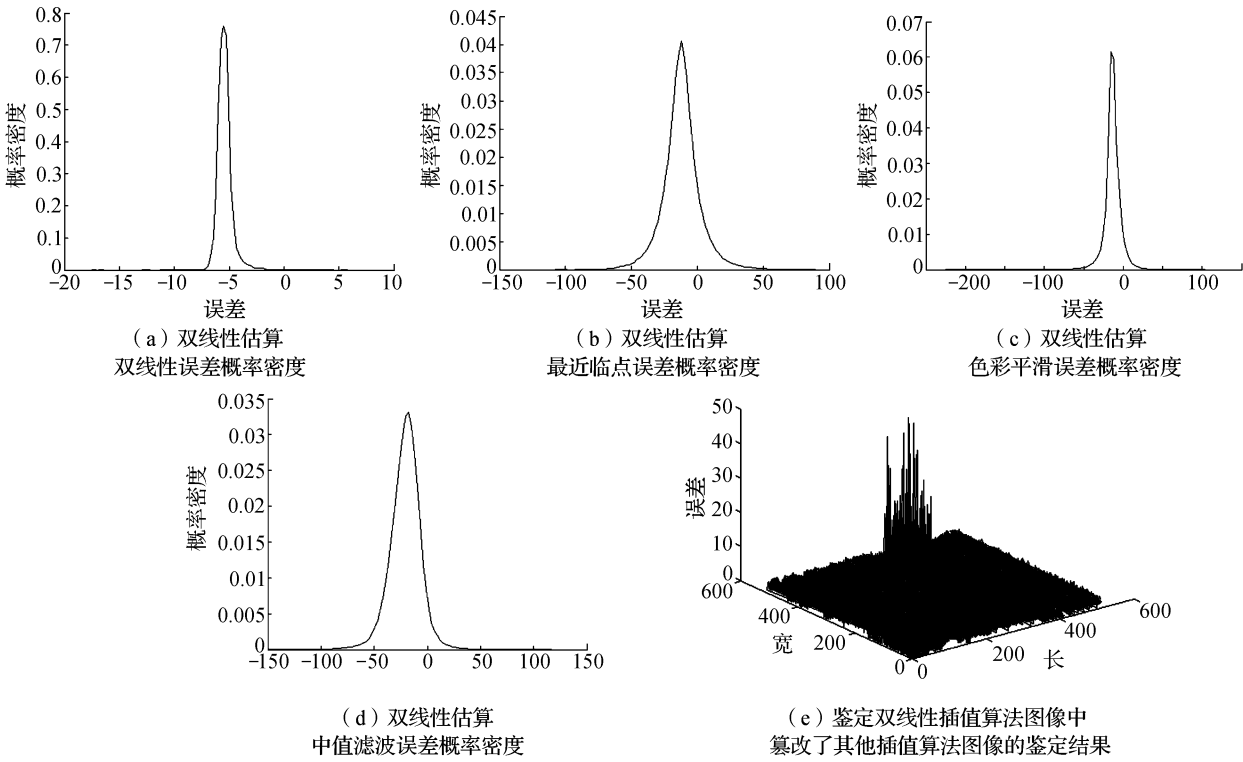


图4 对双线性插值算法图像试验结果

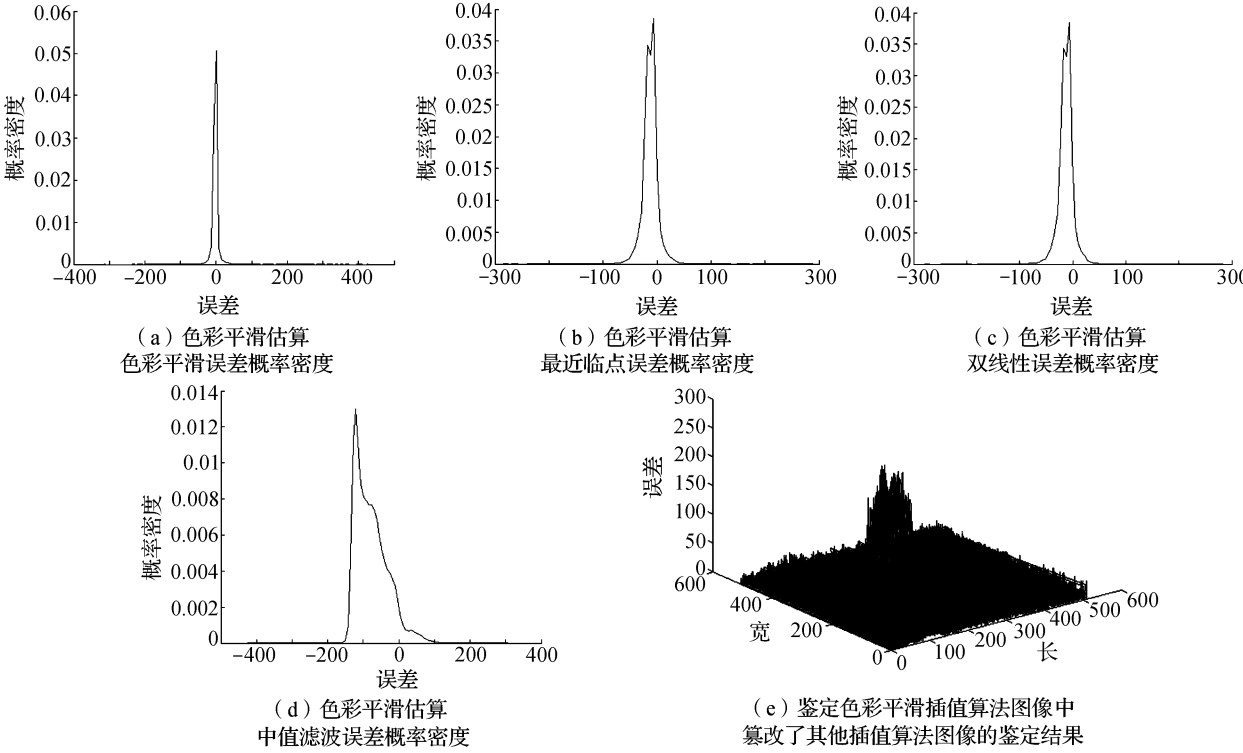


图5 对色彩平滑插值算法图像试验结果

另外本算法对于具有较简单插值关系图像的鉴定效果要明显好于具有较复杂插值算法图像的鉴定效果,即对具有最近临点和双线性插值算法的图像鉴定效果最好,对具有色彩平滑插值算法的图像鉴定效果次之,对具有中值滤波插值算法的图像没有鉴定效果。这是由于BP网络对函数关系的逼近效果与被逼近函数复杂程度成反比的性质所造成的。而中值滤波插值算法在产生图像的过程中,对中间计算参数进行了滤波,更进一步丢失了像素点之间的相互关系,因此从图6(e)中可以看出篡改区域的估算值与实际图像值的误差几乎淹没在整幅图像的估算值与实际图像值的误差当中。

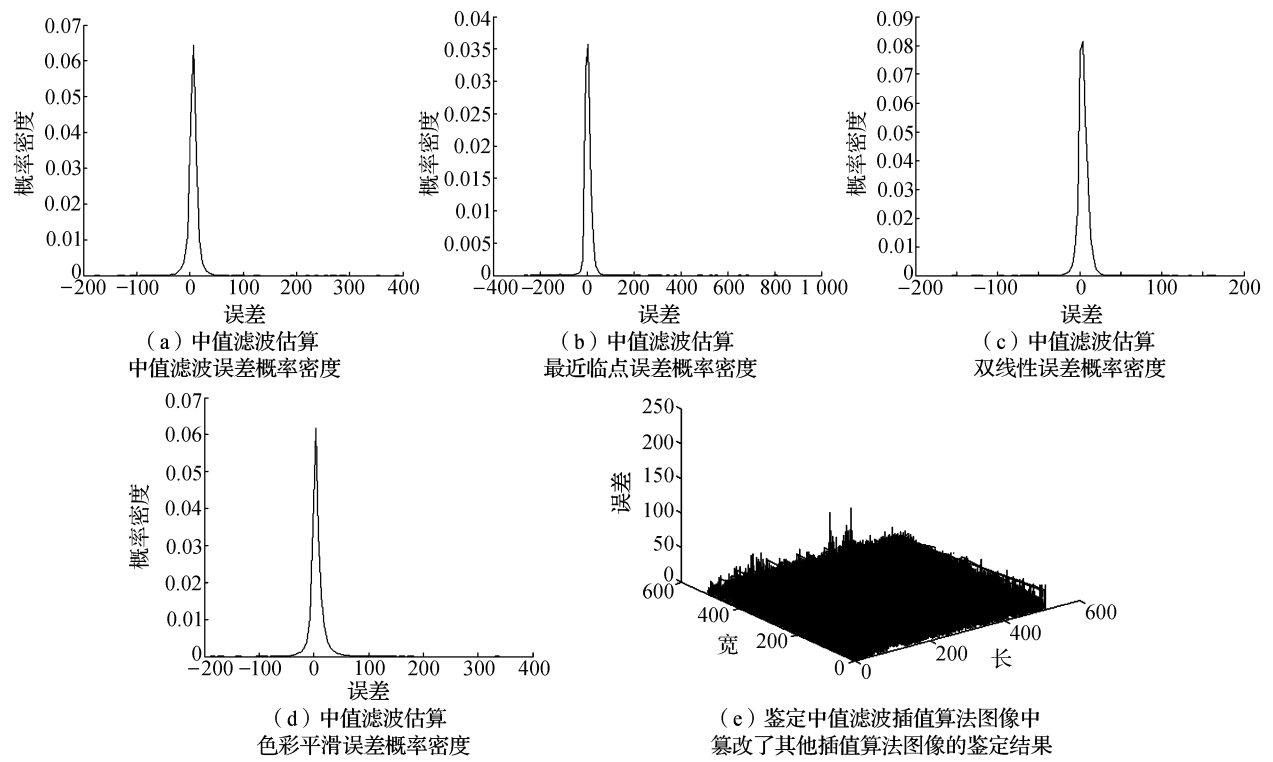


图 6 对中值滤波插值算法图像试验结果

4 结 语

介绍了数码相机的图像插值原理,并对数码相机经常使用的 4 种图像插值算法进行了详细的说明,针对相机的插值特性提出了一种图像盲取证算法,并利用 4 种常用的图像差值算法对本算法进行了系统的实验和分析。实验数据表明,本算法对于由最近临点、双线性、中值滤波 3 种插值算法得到的不含噪声或未使用压缩格式保存的图像有良好的鉴定能力,对采用 JPEG 等压缩格式保存的图像也有很好的鲁棒性。本算法对于由中值滤波插值算法得到的图像不能做出很好的鉴定效果,对复杂的图像有时还存在不收敛的情况,下一步还需要对这两方面问题进行进一步的研究。

参考文献:

[1] 武 伟,詹玲超. 基于数码相机固有特性的篡改检测[J]. 微计算机信息, 2007, 23(10): 131-132.

[2] 李生红,张爱新,郑燕双,等. 基于 SIFT 的图像复制遮盖篡改检测技术[J]. 解放军理工大学学报, 2009, 10(4): 339-343.

[3] 王 波,孔祥维,尤新刚,等. 基于协方差矩阵的 CFA 插值盲检测方法[J]. 电子与信息学报, 2009, 31(5): 1175-1179.

[4] 贺 钦,刘文予. 数字图像传感器颜色插值算法研究[J]. 小型微型计算机系统, 2007, 28(8): 1482-1485.

[5] Popescu A, Farid H. Exposing digital forgeries in color filter array interpolated images[J]. IEEE Transactions on Signal Processing, 2005, 53(12): 3948-3959.

[6] 陈 波,朱秋煜. 基于 EM 算法的人像合成图像的检测[J]. 计算机仿真, 2008, 25(9): 201-203.

[7] Lukac R, Plataniotis K N. Color filter arrays: design and performance analysis[J]. IEEE Transactions on Consumer Electronics, 2005, 51(4): 1260-1267.

[8] 李 萍,曾令可,税安泽,等. 基于 MATLAB 的 BP 神经网络预测系统的设计[J]. 计算机应用与软件, 2008, 25(4): 149-184.