

一种无失真数据隐藏算法^{*}

黄夏菱¹, 刘红梅^{1,2}

(1. 中山大学电子与通信工程系, 广东 广州 510275;

2. 中国科学院研究生院信息安全国家重点实验室, 北京 100039)

摘要: 数字水印和数据隐藏技术有着重要的学术价值和应用前景。但目前现有的绝大多数数据隐藏算法, 都会对原始图像造成不可避免的失真。这在医学应用中是绝对不允许的, 因为引入的失真将可能引起医生对图像的误解。针对数据隐藏在医学影像中的应用, 提出了一种无失真的数据隐藏算法。该算法利用实用的医学图像中存在的冗余, 能够隐藏大量的信息, 而且从嵌入原理上保证了嵌入信息不会对图像的视觉效果有任何损伤。

关键词: 数据隐藏; 医学图像; 数字水印; 无失真

中图分类号: TP391 文献标识码: A 文章编号: 0529-6579 (2002) 06-0018-04

数字水印和数据隐藏技术, 通过在原始媒体数据中嵌入信息来证实该数据的所有权归属或传递信息本身。数字水印和数据隐藏技术有着重要的学术价值和前景, 近年来成为多媒体信号处理领域一个活跃的研究方向^[1,2]。

在典型的数字水印应用中, 隐藏信息与宿主图像间有密切的关系。隐藏信息提供了有关宿主图像的附加信息, 如图像的标题、有关图像来历的辅助数据、作者签名等。隐藏数据在提高图像应用价值的同时, 不可避免地在隐藏过程引入对原始图像的一些失真。一般的数字水印技术在满足一定的稳健性要求以外, 还要求在信息嵌入过程中引入的失真越少越好。人类的视觉系统模型常常用来作为判断由嵌入水印而引起的失真是否达到不可见性的要求。为了尽可能地将数据隐藏过程中引入的失真降到人眼不可觉察的程度, 人们充分地利用人眼的视觉特性, 提出了自适应水印算法^[3-11]。然而, 某些场合却绝对不允许图像有任何的失真, 医学图像就是一个这方面的例子: 哪怕是很微小的改动在法律上都是不允许的, 其潜在的危险是会引起医务人员对图像的错误理解。

在医学应用中, 可用于将病历隐藏在拍摄的医学影像中, 确保病历的秘密性及与诊断依据之间的数据一致性。这在医院是一个相当迫切的需求, 如有些领导人的病历是需要保密的, 而信息隐藏为这

一问题提供了一种很好的方法。但也存在上述的约束条件: 即不允许医学影像有任何失真。

Kodak^[12]在其专利中第一次提出了无失真数据嵌入的概念。Goljan^[13,14]提出了一种可逆的数据隐藏方法, 即允许在嵌入过程中引入失真, 但在检测数据时可以将原始图像恢复出来。其基本思想是将原始图像中用于嵌入数据的特征压缩, 将压缩了的特征和待嵌入数据一起嵌入在原始图像中, 能够嵌入的数据量是特征压缩后的冗余。在数据检测过程中, 由于能够检测到原始图像的特征, 从而可以恢复原始图像。在此文献中, 无失真并不是指图像在嵌入过程中不产生损伤, 而是指这种损伤是可逆的, 即在接收端可以恢复原始图像。但这种方法失去了数据隐藏的一个重要的优点, 即隐藏信息与原始媒体的不可分离性, 一旦隐藏数据从图像中提取出来, 得到的无失真的图像中就不再存在有隐藏数据。

本文针对实用的医学影像图像格式, 提出了一种无失真的数据隐藏方法, 既可以隐藏大量的数据, 又不引起图像的失真, 避免由于数据嵌入引入的失真使观察者对图像产生误解。而且本文提出的方法在保证图像不失真的同时, 使隐藏信息与原始媒体具有不可分离性。

本文第一部分提出算法的数据嵌入原理, 嵌入过程和数据检测过程。第二部分将给出实验结果和

* 收稿日期: 2002-05-01

基金项目: 国家自然科学基金重点资助项目 (60133020); 国家自然科学基金资助项目 (69975011, 60172067); 国家 863 计划资助项目 (2002AA144060); 广东省自然科学基金重点资助项目 (013164); 广东省自然科学基金资助项目 (021758); 中山大学青年教师启动基金资助项目

作者简介: 黄夏菱 (1978 年生), 男, 硕士研究生; 通讯联系人: 刘红梅, E-mail: isslhm@zsu.edu.cn

结论。

1 算法描述

医学中拍摄的某些图像其实际灰度值动态范围较小，可以通过伪彩色进行增强处理，生成伪彩色图像。伪彩色图像数据中包含了调色板（颜色表）和像素值两个部分。通常为了存储的方便，像素值由一个字节表示，像素值是调色板的索引。本节首先介绍伪彩色图像的特征，接着介绍本文提出算法的原理和数据嵌入和检测过程。

1.1 伪彩色图像

如果图像数据中各种颜色的 RGB 值储存在颜色表（color table）或调色板（palette）内，而图像资料中每个像素的颜色值为指向调色板的索引值（index），则该图像称为伪彩色（pseudo color）图像。伪彩色图像是实际应用中较常见的一种医学影像格式。由于图像资料中每个像素的颜色值以不超过一个字节为原则（通常每个像素占 4 个位或 8 个位），因此调色板中最多可以储存 256 种颜色（通常为 16 种或 256 种颜色）。至于图像文件在储存灰度图像（gray scale image）数据时，也是采用 RGB 色彩系统来描述颜色，只不过 R，G，B 三种颜色值都相同罢了。因此，我们可以通过添加调色板，以伪彩色图像的方式来储存灰度图像。

1.2 算法原理

对于伪彩色图像，调色板的每一项对应一种像素值的显示色，如第 0 项对应像素值为 0 时的显示色，第 62 项对应像素值为 62 时的显示色。而现有的算法都是通过改变原始灰度来实现数据的嵌入的。如采用改变像素值最后一个二进制有效位的方法，实际上在嵌入信息时，有可能会改变到水印图像的显示色，如，若原始像素值是 0，但嵌入信息“1”后，实际像素值变成了“1”，则对应的显示色会发生变化，（嵌入“0”，不影响索引色）从而引入了失真。

经过统计，我们发现这一类图像实际的灰度值比较小，通常小于 128。这就意味着在颜色表中的 256 项仅用到很少的一部分，存在着冗余，且原始像素值的最高位是一个冗余位。因此我们提出利用推想中存在的上述冗余嵌入信息。原理如下：

我们将调色板中的 0—127 项拷贝到 128—255 项，用像素值的最高二进制有效位嵌入信息。则在嵌入“0”时，显示时用调色板上半部分的索引色，在嵌入“1”，时，显示时用调色板下半部分索引色。如，原始像素值是 0，嵌入信息位“1”后，

像素值变为 128，显示时的索引色由调色板的第 128 项给出，是原始像素值 0 使用的索引色（第 0 项）的另一个拷贝，因此和没有嵌入信息时是一样的。

利用上述算法，嵌入水印后的显示图像和原始图像完全相同，不会给原始图像带来任何失真，同时又可隐藏大量的信息。

1.3 水印嵌入

水印嵌入过程分为如下：①检测图像；②嵌入水印；③处理（添加）颜色表。

首先，检测原始图像的像素的最大值，如果小于或等于 127（对于 8 bit 图像），则至少有 1 bit 的冗余位（最高位）。

将水印信息嵌入到图像的冗余位（最高位），即直接修改像素值的最高位，如嵌入信息位为“1”，则将最高位置为 1，否则置为 0。

最后，将颜色查找表中第 0—127 项的数据拷贝到第 128—255 项。得到加载水印后的图像。

1.4 水印提取

对于用了本文算法所得到的包含水印的图像，通常只需依次提取像素值的最高位，即可得到所嵌入的信息。

2 实验结果和结论

本文利用给出的算法对一些低像素值的医学图像（spine.tif 490×267×63 colors 和 4 维的大脑切片（128×128×1×27×89 colors）图进行了实验，实验结果见图 1—4。在 spine 中隐藏的数据是长度为 22 478 bytes 的字符串；在 4 维的大脑切片图中第 17 切片 mr 中隐藏的数据是长度为 2 048 bytes 的字符串。为了进行对比，我们采用了 LSB 算法对图像嵌入了同样的信息。改变最后一个位平面的 LSB 算法是一个能很好地保证不可见性，同时隐藏量较大的算法。

图 1 显示了采用本文隐藏算法嵌入信息后的效果。其中图 1（a）是采用本文算法在 spine 图中嵌入长度为 22 478 bytes 的字符串后得到的图像，图 1（b）是含水印图像与原始图像的绝对误差乘以 64 后所得的图像。虽然图中已隐藏了大量的数据，但从图中看不到任何失真的存在。图 2（a）是改变图像像素值最低一个有效位的 LSB 方法在 spine 图中嵌入同样长度为 22 478 bytes 的字符串后得到的图像中，图 2（b）是含水印图像与原始图像的绝对误差乘以 64 后所得的图像。在同样的隐藏量下，可以看到图中明显存在失真。

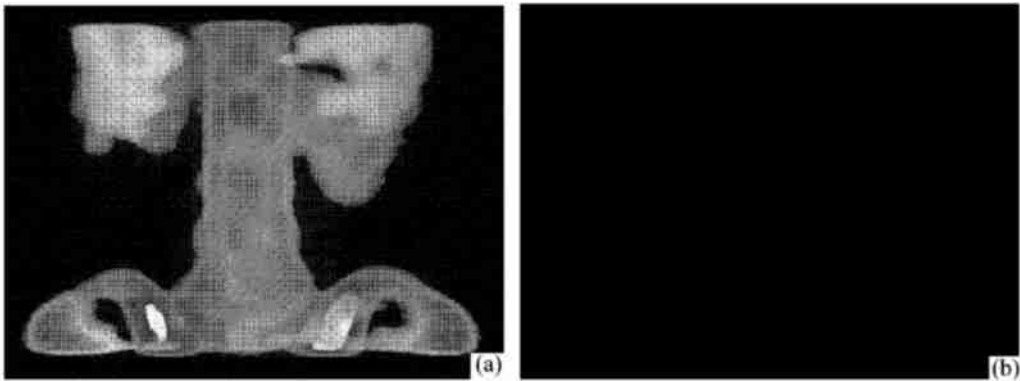


图 1 (a) 采用本文算法处理图 spine 所得含水印图像
(b) 图 (a) 与原始图像绝对误差的放大图

Fig 1 (a) Watermarked "Spine" image processed by the proposed algorithm;
(b) 64 times of the absolute difference against the original image

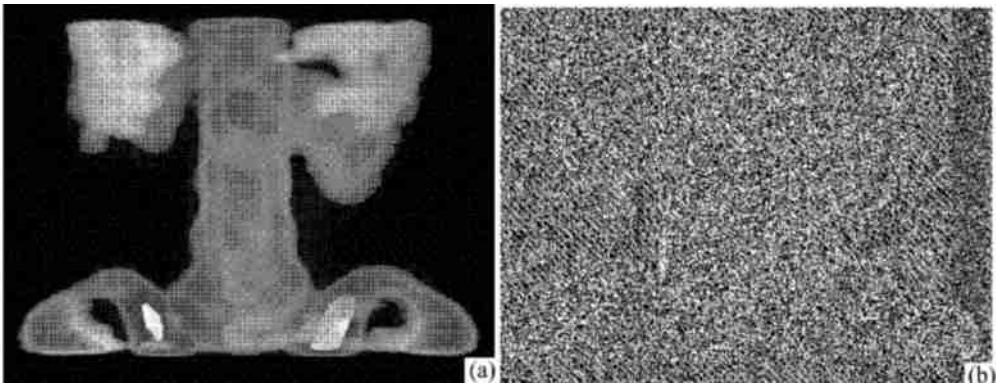


图 2 (a) 采用 LSB 算法处理图 spine 所得含水印图像;
(b) 含水印图像与原始图像绝对误差的放大图

Fig 2 (a) Watermarked "Spine" image processed by the LSB algorithm;
(b) 64 times of the absolute difference against the original image

图 3 是本文提出的无失真数据隐藏方法在图 mr 中嵌入长度为 2 048 bytes 字符串的实验结果。图 3 (a) 是隐藏了数据后得到的图像, 图 3 (b) 是原始图与嵌入图之间的绝对误差的 64 倍, 可以看到然图中虽已隐藏了大量的数据, 但没有任何失真的存在。图 4 是改变图像像素值最低一个有效位的 LSB 算法在 mr 图中嵌入同样长度为 2 048 bytes 字符串的实验结果。图 4 (a) 是隐藏了数据后得到的图像, 图 4 (b) 是原始图与嵌入图之间的绝对误差的 64 倍, 可以看到在同样的隐藏量下嵌入数据的图存在着失真。

由实验结果和算法原理可知, 本文提出的算法具有以下几个方面的贡献:

(1) 采用本文的算法, 嵌入信息后的显示图像与原始图像完全相同, 没有造成任何的失真, 而且可以存储大量的信息。

(2) 本文算法复杂度较低, 且嵌入的容量为已

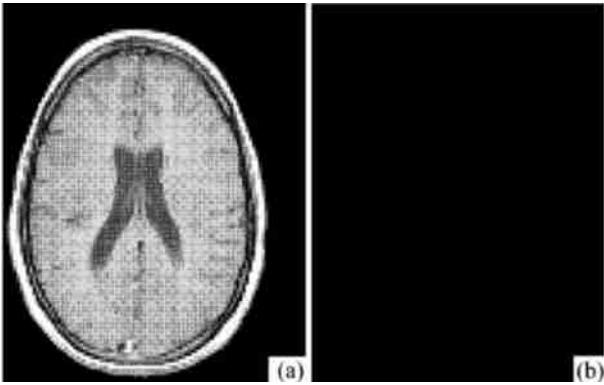


图 3 (a) 采用本文算法处理图 mr 所得第 17 切片;
(b) 图 (a) 与原始切片间的绝对误差的放大图

Fig 3 (a) Watermarked 17th slice in "mr" processed by the proposed algorithm;
(b) 64 times of the absolute difference against the original image

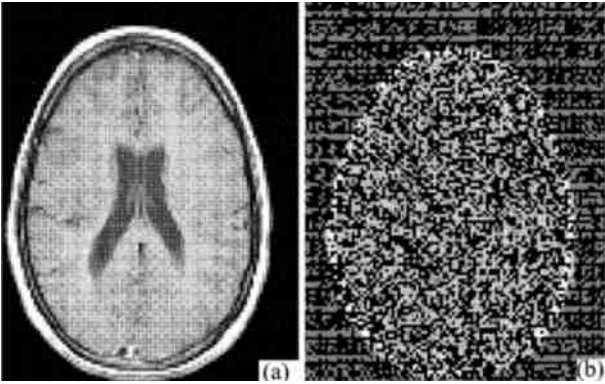


图 4 (a) 采用 LSB 算法处理图 mr 所得第 17 切片;
(b) 图 (a) 与原始切片间绝对误差放大图

Fig 4 (a) Watemarked 17th slice in “mr”
processed by LSB algorithm;
(b) 64 times of the absolute difference
against the original slice

知；而文 [4—6] 的算法复杂度较高，且嵌入容量为不可知，很有可能一幅图不足以嵌入所需的容量。

(3) 本文算法在保证无失真的情况下，保留了数据隐藏的一个重要的优点，即隐藏信息与原始媒体的不可分离性，而文 [4—6] 的算法一旦隐藏数据从图像中提取出来，得到的无失真的图像中就不再存在有隐藏数据，将失去数据隐藏的这个优点。

参考文献：

[1] PETITCOLAS F, ANDERSON R, KUHN M. Information hiding —— a survey[J] . Proc IEEE, 1999, 87(8): 1062—1078.
[2] SWANSON M D, KOBAYASHI M, TEWFIK A H. Multimedia data embedding and watermarking technologies[J] . Proc IEEE, 1998, 86(6): 1064—1087.

[3] 黄继武, 谭铁牛. 图像隐形水印综述[J] . 自动化学报, 2000(5): 645—655.
[4] SWANSON M D, KOBAYASHI M, TEWFIK A H. Multimedia data embedding and watermarking technologies[J] . Proc IEEE, 1998, 86(6): 1064—1087.
[5] SWANSON M D, ZHU B, TEWFIK A H. Multiresolution scene-based video watermarking using perceptual models [J] . IEEE J Select Areas Commun, 1998, 16(4): 540—550.
[6] WOLFGANG R B, PODILCHUK C I, DELP E J. Perceptual watermarks for digital image and video[J] . Proc IEEE, 1999, 87(7): 1108—1126.
[7] HUANG J W, SHI Y Q. An adaptive image watermarking scheme based on visual masking[J] . Electronics Letters, 1998, 34(8): 748—750.
[8] 黄继武, SHI Y Q. 一种自适应图像水印算法[J] . 自动化学报, 1999, 25(4): 476—482.
[9] PODILCHUK C I, ZENG W. Image-adaptive watermarking using visual models[J] . IEEE J Select Areas Commun, 1998, 16(4): 525—539.
[10] WANG H J M, SU P C, KUO C C J. Wavelet based digital image watermarking[J] . Optics Express, 1998, 3(12): 491—496.
[11] DELAIGLE J F, VLEESCHOUWER C De, MACQ B. Watermarking algorithm based on a human visual model[J] . Signal Processing, 1998, 66(3): 319—335.
[12] HONDINGER C W, JONES P, RABBANI M, et al. Lossless recovery of an original image containing embedded data[J] . US Patent Application, Docket No: 77 102/E— D, 1999.
[13] GOLJAN M, FRIDRICH J J, DU R. Distortion-free data embedding for images // Proc of the 4th Int[C] . Workshop on Information Hiding, 2001.
[14] FRIDRICH J, GOLJAN M, DU R. Invertible authentication //Proc SPIE[C] . Security and Watermarking of Multimedia Contents, San Jose, California January, 2001.

A Distortion free Data hiding Algorithm

HUANG Xia ling¹, LIU Hong mei^{1,2}

(1. Department of Electronics, Sun Yat-sen(Zhongshan) University, Guangzhou 510275, China;
2. The State Key Laboratory of Information Security, Beijing 100039, China)

Abstract: Since the digital watermarking and data embedding techniques have academic importance and broad application, they have become one of the active research areas in present multimedia signal processing field. One common drawback of the most of existing data embedding methods is that the original image is inevitably distorted by some small amount of noise. Although the distortion is often quite small and ensured invisible with the models of human visual system, it may not be acceptable for medical images. A distortion free data hiding method is introduced for its special application to medical images. The method can embed large amount of information into the redundant bits of medical images without any visual distortion.

Key words: image watermarking; digital watermarking; data embedding; non distortion