

基于主分量分析的矢量量化数字水印算法

罗 斌, 顾 伟, 吕皖丽, 宫炎焱

(安徽大学计算机科学与技术学院, 合肥 230039)

摘 要: 针对矢量量化过程中码书训练复杂的缺点, 基于主分量分析方法对图像进行降维, 根据降维后各主分量熵和标准差的差异性对其进行分类, 采用非均匀矢量量化方法生成码书。在水印嵌入过程中, 将水印图像嵌入降维后主分量能量适中的码书中以提高水印图像质量, 利用 EENNS 算法代替完全搜索算法缩短编码时间。实验结果表明, 该算法在提高码书质量的同时, 能有效减少码书训练时间, 对 JPEG 压缩、剪切、旋转等图像攻击也具有较强的鲁棒性。

关键词: 主分量分析; 矢量量化; 降维; 码书设计; 数字水印

Vector Quantization Digital Watermark Algorithm Based on Principal Component Analysis

LUO Bin, GU Wei, LV Wan-li, GONG Yan-yan

(School of Computer Science and Technology, Anhui University, Hefei 230039)

【Abstract】 This paper reduces the dimensionality of the images using Principal Component Analysis(PCA) to decrease computation, and the principal components classified by the entropy and the standard deviation are quantized separately. Additionally, the watermark image is embedded in the sub-codebook which the energy of principal components is medium for enhancing images quality. The EENNS algorithm is substitutes for full search algorithm because of falling encoding time. Experimental results show that the proposed algorithm not only improves codebook quality evidently, but also decreases codebook generation time obviously. In addition, it is robust to common image processing operations, such as JPEG compression, cropping, and rotation and so on.

【Key words】 Principal Component Analysis(PCA); Vector Quantization(VQ); dimensionality reduction; codebook design; digital watermark

1 概述

矢量量化(Vector Quantization, VQ)由于其具有压缩率高、解码简单和易于硬件实现的优点, 已广泛应用于图像、语音信号的处理中, 出现了多种基于矢量量化技术的数字水印算法^[1-2]。如何设计性能良好的码书是矢量量化的关键问题之一。设训练矢量集 X 含有 M 个矢量, 码书 CB 的大小为 N 。码书设计就是找出一个最优分类器 Q , 使得在 $Q: X \rightarrow CB$ 中, 量化失真最小^[3]。如基于预测方法的矢量量化算法^[4]、基于模拟退火算法的矢量量化^[5]等, 但这些算法没有降低需要量化的数据量且算法复杂度较高。换言之, 当 M 与 N 相差过大, 即用较少的码字量化数量较多的训练矢量, 不可避免的将产生较大的量化误差, 且 M 与 N 相差越大, 误差越大; M 与 N 相差越小, 则这种误差就越小。针对这些问题, 本文采用主分量分析(Principal Component Analysis, PCA)的方法降低训练矢量集的冗余性和相关性, 减少量化误差; 同时在嵌入水印的过程中, 通过选择合适的水印嵌入位置和嵌入方式提高水印的抗攻击性。

2 基于 PCA 的矢量量化码书设计

2.1 数据降维算法

数据降维算法是将数据由高维空间向低维空间映射得到的数据表示的一种方法, 已广泛的应用于数据分析和图像处理各个领域, 出现了多种降维算法, 如 PCA 算法、ICA 算法等。与其他降维算法不同, PCA 算法在去除图像的二阶相

关性时, 能使重构的图像与原图像相比具有方差小和峰值信噪比高的特点^[6], 主要过程^[7]如下:

设大小为 $s \times t$ 的观测图像 I , 降维后的维数是 q , 且各投影主轴为 a_1, a_2, \dots, a_q , 则根据公式

$$X_j = I \cdot a_j \quad (j = 1, 2, \dots, q) \quad (1)$$

将观测图像 I 降维到各分量 x_1, x_2, \dots, x_q 上。 q 个主分量形成的 $s \times q$ 矩阵 $X = (x_1, x_2, \dots, x_q)$ 称为图像 I 的特征矩阵。令 $A = (a_1, a_2, \dots, a_q)$, 则图像 I 的重建公式为

$$I' = X \cdot A^T = \sum_{j=1}^q x_j \alpha_j^T \quad (2)$$

2.2 改进的矢量量化码书设计

降维后的数据量将大量的减少, 如果将降维后数据直接进行矢量量化, 虽然算法的运行时间会比传统的 LBG 算法运行时间要短, 但算法的性能会受到影响, 使得重构后的图像质量较低。针对这一问题, 本文提出自己的改进。

传统的码书设计算法是将所有的训练矢量作为一个整体进行训练, 生成最终的码书。但经过 PCA 降维后的数据差异较大, 能量主要集中在前几个主分量中。若将所有的主分量

基金项目: 国家自然科学基金资助项目(60772122); 博士点基金资助项目(20070357001)

作者简介: 罗 斌(1963 -), 男, 教授, 主研方向: 图像处理, 模式识别; 顾 伟, 硕士研究生; 吕皖丽, 讲师; 宫炎焱, 硕士研究生
收稿日期: 2009-06-10 **E-mail:** maths2007@163.com

放在一起量化则不可避免的产生较大的误差。因此，本文将降维后的数据进行非均匀的分类量化，通过计算熵和标准差两个特征，比较各主分量之间的差异，将差异较小的主分量作为一类，然后再对每类数据分别加以量化，使不同类的主分量采用不同的步长量化，从而降低量化过程中的误差。主要步骤如下：

Step1 设观测图像 I 的大小为 $s \times t$ ，输入参数 τ ，利用二维 PCA 算法对其进行降维，得到主分量 x_1, x_2, \dots, x_q ，令 $X=(x_1, x_2, \dots, x_q)$ ；

Step2 将各主分量 x_1, x_2, \dots, x_q 分别进行数据处理，按映射 f 将数据映射至 $0, 1, \dots, M$ 的范围，即 $f: x_i \rightarrow x_i'$ 。这里 M 取 X 的最大值；

Step3 计算各主分量 x_1', x_2', \dots, x_q' 的熵和标准差，将熵和标准差较为相近的主分量作为一类 cl_i ，令 $CL=\{cl_1, cl_2, \dots, cl_r\}$ ；

Step4 对每类主分量集合 cl_i 分别采用 LBG 算法进行矢量量化，得到码书 $cb_i' (i=1, 2, \dots, r)$ ；

Step5 按映射 f^{-1} ，将码书 cb_i' 还原为 cb_i ，合并 cb_i ，生成码书 $CB=\{cb_1, cb_2, \dots, cb_r\}$ 。

3 数字水印的嵌入和提取算法

3.1 水印的嵌入算法

为提高嵌入水印后图像的质量，本文选择将水印信息嵌入于能量适中的主分量所对应的码书中，舍弃能量较大或较小的主分量所对应的码书。因为如果将水印嵌入于能量较大的主分量中，则产生较大的误差，影响图像质量；如果将水印嵌入于能量较小的主分量中，水印信息容易被各种图像攻击算法所滤除。同时由于矢量量化编码过程是影响算法实时性的一个重要因素，为缩短编码时间，提高算法实用性，本文采用 EENNS 搜索算法^[8]代替传统的完全码字搜索算法。最后，本文采用差分曼彻斯特编码的方式嵌入水印，增强水印的鲁棒性。

设载体图像为 I ，水印图像为 W 。其中， $I=\{I_{ij} | 1 \leq i \leq s, 1 \leq j \leq t\}$ ， I_{ij} 代表 (i, j) 处的像素灰度值。 $W=\{w_{ij} | 1 \leq i \leq m, 1 \leq j \leq n\}$ ， $w_{ij} \in \{0, 1\}$ 。

水印嵌入过程如下：

Step1 将水印图像 W 进行 Arnold 置乱，并将置乱后的水印图像按列展开为 1 行 k 列的水印序列 $w=\{\omega_0, \omega_1, \dots, \omega_k\}$ 。其中， $k=(m \times n)$ 。

Step2 采用二维 PCA 算法将载体图像 I 降维，并生成码书 CB 。

Step3 比较降维后各主分量 x_1, x_2, \dots, x_q 的能量，选择能量适中的主分量 y_1, y_2, \dots, y_h 所对应的码书 CB' ，将其均分为 2 组 P_0, P_1 ，使 2 组中对应位置码字的欧氏距离最短。

Step4 采用 EENNS 算法在码书 CB' 中搜索与各主分量 y_1, y_2, \dots, y_h 距离最近的码字 $c_i (i=1, 2, \dots, k)$ ，同时依次读入水印序列 w 的值，若 $\omega_i=0$ ，则要求与主分量相对应的码字 $c_i \in P_0$ ，如果 $c_i \in P_1$ ，则采用与 P_1 相对应的属于 P_0 的码字代替；同理，若 $\omega_i=1$ ，则要求 $c_i \in P_1$ 。并依次记录码字 c_i 与各个主分量之间的对应关系 I_{c1} 。

Step5 将其余主分量也采用 EENNS 算法，在码书 CB 中找出与之距离最近的码字，记录二者对应关系 I_{c2} ，生成 $I_c=\{I_{c1}, I_{c2}\}$ 。

Step6 根据对应关系 I_c ，使用码书中各个码字代替各主

分量，并根据式(2)生成水印图像 I' 。

3.2 水印的提取算法

水印提取算法本质上是在码书中寻找与各主分量距离最近的码字，并根据水印嵌入规则提取水印的过程，算法计算量小、时间复杂度低。水印提取时，只需使用相同的码书 CB 即可提取水印图像，因此，本水印算法属盲水印提取算法。步骤如下：

Step1 依据 $X=I' \cdot A$ ，从含有水印的图像 I' 还原出主分量矩阵 X ，同时依据各主分量能量的大小，找出能量适中的主分量 y_1', y_2', \dots, y_h' 。

Step2 将码书 CB' 按相同的方式进行分组，生成 P_0, P_1 。

Step3 在码书 CB' 中依次查找与各主分量 y_1', y_2', \dots, y_h' 距离最近的码字 c 。若 $c \in P_0$ ，则记 $w_i'=0$ ；若 $c \in P_1$ ，记 $w_i'=1$ 。

Step4 将序列 $w'=\{w_0', w_1', \dots, w_k'\}$ 还原成大小为 $(m \times n)$ 的矩阵，并进行 Arnold 置乱，产生水印图像。

4 实验结果与分析

实验选用 256×256 大小的 256 灰度级 Lena, Pepper, Woman, Camera 图像作为载体图像、 32×32 大小的二值图像作为水印图像。采用归一化相关值 NC ^[9] 评价提取的水印效果；采用峰值信噪比 $PSNR$ 度量嵌入水印后的图像质量^[10]。

$$PSNR = 10 \lg \left(\frac{255^2}{MSE} \right) \text{ dB} \quad (3)$$

$$MSE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (X(i, j) - X'(i, j))^2 \quad (4)$$

$$NC = \frac{\sum_{j=1}^m \sum_{i=1}^n [(\omega_{ij} - A) \times (\omega'_{ij} - \hat{A})]}{\sqrt{\sum_{j=1}^m \sum_{i=1}^n (\omega_{ij} - A)^2} \times \sqrt{\sum_{j=1}^m \sum_{i=1}^n (\omega'_{ij} - \hat{A})^2}} \quad (5)$$

其中， $A = \frac{\sum_{j=1}^m \sum_{i=1}^n \omega_{ij}}{m \times n}$ ， $\hat{A} = \frac{\sum_{j=1}^m \sum_{i=1}^n \omega'_{ij}}{m \times n}$ 。

(1) 降维后的维数与图像质量的关系

以 Lena 图像为例，图 1 比较了使用 PCA 算法和 ICA 算法进行降维和量化后，维数 q 与未嵌入水印的图像 $PSNR$ 值之间的关系。可以看出，在不同的降维方法中由 PCA 算法得到码书质量较高，适合矢量量化后嵌入水印信息。

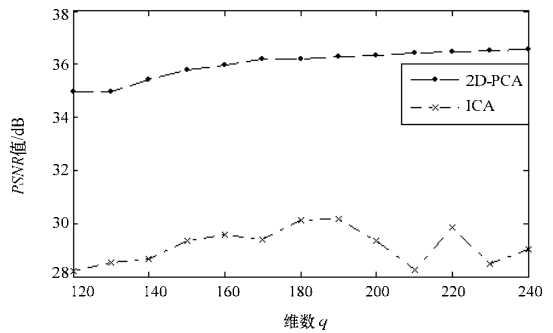


图 1 采用不同降维方法 Lena 图像 $PSNR$ 值随维数 q 的变化

图 2 表示将各载体图像分别采用 PCA 算法降维和量化后，维数 q 与未嵌入水印的图像 $PSNR$ 值之间的关系。可见，图像质量随维数 q 的不断减少而降低，且当维数在 180 左右变化时，图像 $PSNR$ 值会有较明显的变化。换言之，当维数大于 180 时，图像保留了较多的特征信息，因而维数的变化对图像质量影响较小；而当维数过小时，图像将损失过多的细节特征，所以图像质量影响较大，造成 $PSNR$ 值的快速下降。

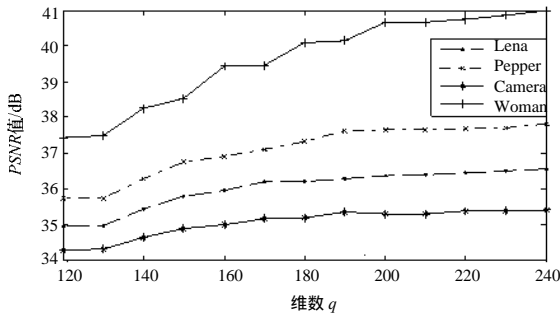


图2 图像 PSNR 随维数 q 的变化

(2) 码书质量及码书生成时间的比较

表1 比较了当维数为 180 时, 在码书大小不同的条件下使用 PCA 算法降维后并分类量化得到的嵌入水印的图像 PSNR 值。可以看出, 码书大小与图像质量是正相关的。在相同的条件下, 码书越大, 则图像 PSNR 值越高, 图像质量越好。图3 显示了码书大小为 1 024 时嵌入水印的 Lena 图像和 Woman 图像。

表1 码书大小不同的条件下各载体图像的 PSNR 值

码书大小	Lena/dB	Pepper/dB	Camera/dB	Woman/dB
256	31.556 2	32.406 9	31.279 3	34.254 9
512	32.280 6	33.198 5	31.944 8	35.52 4
1 024	33.747 2	34.616 9	32.826 5	37.37 5



(a)Lena 图像



(b)Woman 图像

图3 码书大小为 1 024 时嵌入水印的图像

文献[10]采用遗传算法对经过 PCA 降维后的数据进行优化得到码书。文献[11]采用 LBG 算法生成码书并对码书进行处理从而嵌入水印。以 Lena 图像为例, 表2 分别列出了将图像采用 PCA 算法和 ICA 算法降维后进行分类量化并嵌入水印的图像 PSNR 值(以下简称 PCA 算法+分类量化、ICA 算法+分类量化)、采用文献[10]算法并嵌入水印后的图像 PSNR 值(以下简称 PCA 算法+GA 算法)以及采用文献[11]方法得到的含水印的图像 PSNR 值。可以看出, 由本文算法所得到的码书质量要优于其他方法, 这主要因为采用 PCA 算法降维得到的各主分量能量大小差异较大, 而本文采用的分类量化方法正好能够利用主分量能量差异性的特点对其进行分类, 从而有利于提高矢量量化的结果。

表2 不同算法图像质量的比较

码书大小	PCA 算法+GA 算法/dB	文献[11]算法/dB	ICA 算法+分类量化/dB	PCA 算法+分类量化/dB
128	27.48	29.43	26.12	31.14
256	28.74	30.97	27.35	31.56
512	31.54	32.00	28.02	32.28
1 024	32.24	32.62	28.26	33.75

表3 将 LBG 算法、PCA+GA 算法、ICA+分类量化算法以及 PCA+分类量化算法的码书生成时间在 Pentium IV 计算机上进行比较, 并以 Matlab 为平台实现。可以看出, 本文所采用的算法生成码书的时间最短, 远小于 LBG 算法和 PCA+GA 算法所需时间。这主要因为 LBG 算法和文献[12]算法实质是一个反复迭代的运算过程, 占用大量时间; 而本

文算法是先将降维后的主分量进行分类再生成码书, 可减少大量时间。

表3 码书生成时间比较

码书大小	LBG 算法/s	PCA 算法+GA 算法/s	ICA 算法+分类量化/s	PCA 算法+分类量化/s
128	47.88	10.71	2.21	3.85
256	70.96	11.28	5.65	5.87
512	148.89	38.45	11.42	8.68
1 024	267.43	56.78	19.90	13.33

(3) 水印鲁棒性测试

表4 列出了将 Lena 图像使用 Photoshop 7.0 进行图像处理采用 PCA+分类量化算法和文献[11]方法提取水印的实验结果。由表可知, 本文算法对 JPEG 压缩具有较强的抗压缩能力; 对图像剪切攻击, 即使在 1/2 剪切的情况下, 水印 NC 值仍能达到 70% 以上; 但椒盐噪声、滤波等其他图像攻击对水印的提取有较大的影响, 这是因为当采用这些方法对图像进行攻击时, 将使图像的各主分量有较大变化, 从而影响水印的提取效果。

表4 将 Lena 图像进行处理后提取的 NC 值

图像处理	文献[11]方法	PCA 算法+分类量化
JPEG 压缩(质量品质 90)	0.998 2	1.000 0
JPEG 压缩(质量品质 60)	0.977 9	0.982 0
剪切 1/4	0.857 2	0.864 2
剪切 1/2	0.678 4	0.702 6
椒盐噪声(方差 0.01)	0.993 1	0.964 9
高斯低通滤波	0.997 2	0.841 4
旋转 0.1°	0.998 2	0.998 3
锐化	0.996 5	0.970 3
高斯模糊	0.996 8	0.812 5

5 结束语

矢量量化技术是一种有效的有损压缩技术, 具有压缩比高以及解码简单的优点。本文将图像降维的方法引入码书设计中, 针对降维后不同的分类数据分别加以矢量量化, 提高码书质量; 同时选取合适的水印嵌入位置和方式, 增强水印的抗攻击能力。

但本文对码书的设计和优化是在空域中实现的。如何将变换域数字水印算法的优点和矢量量化数字水印技术相结合是今后研究的重点, 以期能得到更好的数字水印算法。

参考文献

- [1] Wang Feng-Hsing, Lakhmi C J, Jeng-Shyang Pan. VQ-based Watermarking Scheme with Genetic Codebook Partition[J]. Journal of Network and Computer Applications, 2007, 3(8): 304-323.
- [2] Luo Bin, Gu Wei, Guo Hui. A VQ Digital Watermark Algorithm Based on T-mixture Models Segmentation[C]//Proc. of IEEE Conference on Neural Networks and Signal Processing. Zhengjiang, China: IEEE Press, 2008: 353-358.
- [3] Shen Furoo, Hasegawa O. An Adaptive Incremental LBG for Vector Quantization[J]. Neural Networks, 2006, 19(5): 694-704.
- [4] Li Yuenan, Liu Chunhe, Lu Zheming. Robust Image Watermarking Algorithm Based on Predictive Vector Quantization[C]//Proc. of the 1st International Conference on Innovative Computing, Information and Control. Beijing, China: [s. n.], 2006.
- [5] 王可, 王翠梅. 用模拟退火算法实现语音识别中的矢量量化[J]. 电子与信息学报, 2000, 22(1): 19-22.
- [6] 冯燕, 何明一, 宋江红, 等. 基于独立分析的高光谱图像数据降维及压缩[J]. 电子与信息学报, 2007, 29(12): 2871-2875.

(下转第 172 页)