

基于块边缘模式的图像内容描述符^{*}

康 勤 邱开金 肖国强

(西南大学计算机与信息科学学院 重庆 400715)

摘要 本文提出一种有效的基于从图像压缩域提取的块边缘模式的内容描述符,它特别适用于 JPEG 压缩图像或 MPEG 压缩视频。该图像内容描述符由三种块边缘模式的游程编码直方图构成。大量的实验结果证明,与现行的类似算法相比,本文提出的算法在进行图像检索时,具有更高的平均查准率和查全率。

关键词 JPEG, MPEG, 图像内容描述符, 图像检索

Block-edge-pattern Based Image Content Descriptor

KANG Qin QIU Kai-jin XIAO Guo-qiang

(Faculty of Computer & Information Science, Southwest University, Chongqing 400715, China)

Abstract In this paper, we describe a robust and effective content descriptor based on block edge patterns extracted directly in DCT domain, which is suitable for applications in JPEG or MPEG compressed images and videos. This content descriptor is constructed by an edge-block histogram with three patterns of horizontal edge, vertical edge and no-edge. Extensive experiments also support that the proposed content descriptor is effective in representing and describing image content. Compared with comparable existing technique, the proposed descriptor achieves superior performances in terms of retrieval precision and recall rates.

Keywords JPEG, MPEG, Content descriptor, Image retrieval

1 引言

目前大多数分布在 Internet 或存储于计算机里的视频或图像都是以压缩的格式存在,因此,在压缩域内对图像或视频进行处理就变得越来越重要。为了从压缩的图像或视频里检测或提取内容特征,传统的方法需要首先把图像解码到像素域,然后才能进行其它的处理^[1-6]。为了避免解码所带来的额外计算代价,目前许多研究都直接从压缩域中提取图像的特征,如纹理、色彩和形状等。在相关领域的代表性工作概括如下。

Shneier 等^[18]通过计算一个窗口内 DCT 系数的平均值来产生 JPEG 图像的特征,从而进行基于内容的图像检索。S. F. Chang 等^[1]计算 DCT 系数的统计测度以构成纹理特征。Ngo 等^[2]提出一种图像检索算法,它通过重新组织 DCT 系数,从而在压缩域内表示色彩、形状和纹理等特征。G. C. Feng 和 J. Jiang^[3]直接从 DCT 系数中提取均值和标准差作为统计特征来进行 JPEG 压缩图像的检索。文献^[4]报道了一种在 DCT 域的快速的纹理描述和图像检索的算法。Liu 等^[5]利用小波系数来进行图像检索。Zhong 等^[6]提出一种基于量化 DCT 块的直方图方法,这些直方图可通过对量化因子和 DCT 块的数量进行优化,进而获得最好的图像检索性能。

在过去几十年里,对图像处理的大量研究表明,图像的边缘是刻画图像内容的一个重要特征,在帮助图像的内容理解上起到较大的作用。如上所述的许多方法都是在压缩域内提取图像的特征,然而目前大多数文献报道的图像边缘提取还是在像素域里进行,例如被广泛应用的图像分割技术^[7]。

近年来,在像素域或压缩域中,以宏块级的图像边缘检测和边缘模式分类成为图像处理的一种趋势^[8-9]。Kim 等^[10]发现 DCT 系数的投影极性能很好地表示边缘位置。Soltane 等^[7]提出一个基于 DCT 系数的均值、方差和熵的边缘选择方案来进行图像分割。Shen 等^[11]通过对 DCT 系数的模式分析,进一步提出一种确定边缘强度和方向的方法。他们通过对理想边缘模型的 DCT 系数的分析,发现可以在宏块级确定不同 DCT 系数的相对值和符号。

最近,由 Chang 等^[12]提出了一种快速地提取 DCT 域中每个块的边缘模式的方案,它把块边缘模式分为五种:无边、水平边缘、45°边缘、垂直边缘、135°边缘,并且提取这五种块边缘模式的计算复杂度也较低。他们的方法与文献^[13]提出的方法类似。

通过对上述的相关工作的分析,本文提出一种基于块边缘模式的图像内容描述符,它直接从 DCT 域中提取三种块边缘模式,分别为无边、水平边缘、垂直边缘。通过三种块边缘模式的不同组合,并形成相应的直方图(我们提出 3-D、4-D 和 run-length 直方图)来表示压缩图像或视频的内容。

2 块边缘模式的提取

2005 年,Chang 等人^[12]提出一种从 8×8 像素块的 DCT 系数提取五种块边缘模式的算法。该算法仅需要分析前几个 DCT 系数,因此不需要完全译码。给定一个 8×8 像素块,把它分成 4 个 4×4 的子像素块,从左到右,从上到下的子像素块的平均像素值分别表示为 S_{00} , S_{01} , S_{10} , S_{11} 。利用表 1 分别计算出五种边缘模式的度量值,由其最大值确定该像素块

^{*} 重庆市自然科学基金(CSTC-2006BB2309)资助项目。康 勤 工程师,研究方向为数字化资源;邱开金 讲师,研究方向为图像处理;肖国强 教授,硕士生导师,研究方向为数字媒体处理。

的块边缘模式。

表 1 五种边缘模式的度量值计算

Edges	Measure values
No edge	$\delta_{NE}(\text{set by user})$
0	$\delta_0 = (S_{00} + S_{01})/2 - (S_{10} + S_{11})/2 $ $\delta_{\pi/4} = \max\{ S_{00} - (S_{01} + S_{10} + S_{11})/3 , S_{11} - (S_{00} + S_{01} + S_{10})/3 $
$\pi/4$	
$\pi/2$	$\delta_0 = (S_{00} + S_{10})/2 - (S_{01} + S_{11})/2$
$3\pi/4$	$\delta_{3\pi/4} = \max\{ S_{01} - (S_{00} + S_{10} + S_{11})/3 , S_{10} - (S_{00} + S_{01} + S_{11})/3 $

经过研究发现,在不影响表征图像内容的前提下,采用三种块边缘模式(即无边缘、水平边缘、垂直边缘)来重构图像的边缘,对图像边缘的表征并不会产生明显的影响,这样可以减少块边缘模式数目,降低计算复杂度。为了论证我们提出的观点,我们重建一个块边缘模式图像,并计算其相对于原图像的峰值信噪比(PSNR)^[17]。由于五种块边缘模式中已包含三种块边缘模式,因此我们仅需定义五种块边缘模式图像的像素值,如表 2 所示。计算出 6 个样本图像对应的 PSNR 值如表 3 所示。

表 2 定义五种块边缘模式图像的像素值

Edges	$\overline{p_{ij}}$
No edge	$\overline{p_{ij}} = DC$
0	$\overline{p_{ij}} = (S_{00} + S_{01})/2, j \in \text{two sub_blocks at the top};$ $\overline{p_{ij}} = (S_{10} + S_{11})/2, j \in \text{two sub_blocks at the bottom};$
$\pi/2$	$\overline{p_{ij}} = (S_{00} + S_{10})/2, j \in \text{two sub_blocks on the left};$ $\overline{p_{ij}} = (S_{01} + S_{11})/2, j \in \text{two sub_blocks on the right};$
$\pi/4$	if $\delta_{\pi/4} = S_{00} - (S_{01} + S_{10} + S_{11})/3;$ $\overline{p_{ij}} = S_{00}, j \in \text{top left sub_block}; \overline{p_{ij}} = (S_{01} + S_{10} + S_{11})/3, i, j \in \text{all other sub_blocks};$ if $\delta_{\pi/4} = S_{11} - (S_{01} + S_{10} + S_{00})/3;$ $\overline{p_{ij}} = S_{11}, i, j \in \text{bottom right sub_block}; \overline{p_{ij}} = (S_{00} + S_{01} + S_{10})/3, i, j \in \text{all other sub_blocks};$
$3\pi/4$	if $\delta_{3\pi/4} = S_{01} - (S_{00} + S_{10} + S_{11})/3;$ $\overline{p_{ij}} = S_{01}, i, j \in \text{top right sub_block}; \overline{p_{ij}} = (S_{00} + S_{10} + S_{11})/3, i, j \in \text{all other sub_blocks};$ if $\delta_{3\pi/4} = S_{10} - (S_{01} + S_{11} + S_{00})/3;$ $\overline{p_{ij}} = S_{10}, j \in \text{bottom left sub_block}; \overline{p_{ij}} = (S_{00} + S_{01} + S_{11})/3, i, j \in \text{all other sub_blocks};$

表 3 边缘模式图像相对于原图像的 PSNR

Image Samples	PR ₅ (dB)	PR ₃ (dB)
Lena	59.439892	59.180862
Peppers	57.447807	57.687534
Clown	50.144257	50.109631
Camera	48.590816	48.425625
Boat	55.761734	56.058308
Bridge	47.100334	47.334343

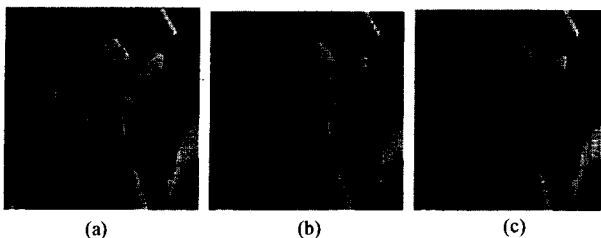


图 1 (a) Lena 原图 (b) 五种边缘模式图像 (c) 三种边缘模式图像

从表 3 的数据可知,分别用三种块边缘模式和五种块边缘模式表示图像时,其图像的 PSNR 相差甚微。由此可见,利用文献 [12] 的边缘模式提取方法进一步构建内容描述符时,仅需要水平边缘和垂直边缘即可,从而大大降低了计算量,便于实时应用和高速处理。图 1 给出了 Lena 图像在不同块边缘模式下的图像,给我们的视觉感知仅带来微小的差别。

根据文献[12,13],两种块边缘模式的度量值可直接从 DCT 系数获得,即:

$$\delta_0 = |X(1,0)| \quad (1)$$

$$\delta_{\pi/2} = |X(0,1)| \quad (2)$$

其中, $X(1,0)$ 和 $X(0,1)$ 分别是沿 zig-zag 扫描的第 2 和第 3 个 DCT 系数。因此,对于给定的像素块,它的块边缘模式可由下式确定:

$$\text{block_edge_pattern} = \begin{cases} \text{no_edge} & \text{if } \max(\delta_0, \delta_{\pi/2}) < \lambda \\ \text{vertical_edge} & \text{if } \delta_{\pi/2} \geq \delta_0 \\ \text{horizontal_edge} & \text{otherwise} \end{cases} \quad (3)$$

其中, λ 为用实验方法确定的门限,本文中 $\lambda = 50$ 。

3 图像内容描述符

给定一个 8×8 的像素块,它们的块边缘模式用 3 个元素的集合 $\psi = \{\#, -, |\}$ 表示,其中,符号“#”、“-”和“|”分别表示无边缘模式、水平边缘模式和垂直边缘模式。如果以每种块边缘模式出现的次数来构造直方图(1 维块边缘直方图, 1D-EBH),则该直方图只有 3 个柄(bin)。由于 1D-EBH 的柄太少,难以反映图像的内容,因此,我们把相邻的块边缘模式进行组合,以便构造高维直方图。例如,把两个相邻的块边缘模式进行组合作为 1 个状态来考虑,则由此构造的直方图的柄可增加到 9 个(即 2D-EBH);若把 N 个相邻的块边缘模式组合作为 1 个状态来考虑,则总的直方图的柄可增加到 3^N 个。

低维直方图具有较少的柄,但它所包含的块边缘在图像中的位置信息也较少,而这些位置信息对反映图像内容是极其重要的。另一方面,高维直方图(如 5D-EBH 和 6D-EBH)虽然包含较多的位置信息,但它的块边缘模式组合后的状态数目较大,从而计算量很大,增加了算法的计算复杂度。为了平衡两者之间的关系,我们受游程编码思想的启发,提出一种构造游程块边缘直方图(RLEBH)的方法。对一幅图像分别按行和列的顺序进行扫描,并统计相同块边缘模式在一个游程内出现的次数。当出现一个新的块边缘模式时,认为当前游程结束,一个新的游程开始。每个游程的输出编码表示为:

$$R_i = \{\alpha_i, \eta_i\} \quad (4)$$

其中, $\alpha_i \in \{\#, -, |\}$, η_i 表示一个游程中 α_i 出现的次数。

R_i 代表 RLEBH 的状态(即柄)。为了降低统计模型的复杂度^[14],把游程长度大于阈值 T 的归为同一状态,即:

$$R_i = \begin{cases} \{\alpha_i, \eta_i\} & \text{if } \eta_i \leq T \\ \{\alpha_i, T\} & \text{when } \eta_i > T \end{cases} \quad (5)$$

实验中 $T = 4$ 。因此 RLEBH 可表示为:

$$H_{run} = \frac{\{C_1, C_2, \dots, C_M\}}{M} \quad (6)$$

其中, C_i 表示对一幅图像进行游程编码后 R_i 出现的次数, $M = 24$, 包括 12 个基于行扫描的游程状态和 12 个基于列扫描的游程状态。

式(6)给出的直方图不仅解决了高维直方图带来的直方

图的柄随维数成指数增长的问题,而且提供了块边缘模式分布的非冗余统计分析。这主要得益于游程编码本身具有的特性,它把连续的不同块边缘模式作为一个状态来对待,而并非对每个块边缘模式孤立地作为一个状态来考虑。同时在所得到的 RLEBH 中,在一个游程内的块边缘模式还能够反映相邻边缘的关系,这样,我们提出的 RLEBH 不仅能够描述在块边缘模式分布意义上的视觉内容,而且这些连接的边缘能够提供物体的形状、轮廓和纹理等信息。我们提出的内容描述符不同于文献[8,9]的基于边缘的描述符。文献[8]中的描述符依赖于连接 Canny 算子提取的突点,以形成边缘曲线。文献[9]是首先提取边缘图,然后通过模糊聚类和形状信息的组合来形成全局特征,用于图像检索。

另外,本文提出的 RLEBH 能够保留传统直方图的强度,并包含有图像的边缘连通性和区域块边缘模式的连续性信息。因为游程直方图是分别由沿行和列扫描一幅图像的块边缘后构成的,所以,RLEBH 具有 2 维边缘连通性信息,而不是 1 维的。因此,其全局连通性(如对角边缘连通性)可认为是由基于行和列的边缘连通性共同起作用的结果。

4 实验结果与分析

为了对本文提出的内容描述符进行评估,我们设计了两个阶段的实验。第一阶段的实验是用所提出的内容描述符进行图像检索,并用查准率(precision)来衡量其效果。为此我们建立了一个由 14787 幅 JPEG 图像组成的图像库,并用人工的方式把图像分成 10 类,包括汽车、花卉、自行车、建筑物、动物、人物、风景、天空、日落和纹理图像,以便为实验提供一个比对的基准。第二阶段的实验,我们随机从图像库中选取图像,并对其进行一系列的方位变换和加噪处理后,再放回图像库中,用原图像作为查询图像在图像库中进行检索,根据检索出的经上述变化的图像,计算其查全率(recall),第二阶段的实验主要是验证本文提出的内容描述符的稳健性。

为了与现行的算法进行比较,我们以文献[3,15]的算法作为 benchmark。第一个 benchmark,即文献[3]的算法是在压缩域内直接提取 8×8 像素块的均值和方差,把均值和标准差分别量化成 4 段和 7 段,然后构造直方图。因此,该算法与本文提出的算法有很好的可比性。第二个 benchmark,即文献[15],它首先通过下式对 RGB 彩色空间变换为恒定彩色空间(invariant color space)

$$\{c_4, c_5, c_6\} = \begin{cases} c_4 \{RGB\} = \frac{R-G}{R+G} \\ c_5 \{RGB\} = \frac{R-B}{R+B} \\ c_6 \{RGB\} = \frac{G-B}{G+B} \end{cases} \quad (7)$$

并把恒定彩色空间量化成 4 个等级,然后构造直方图,以此来对图像进行检索。

为了计算直方图之间的相似度,我们用 L_1 距离,因为:(1)它广泛应用于 CBIR 系统,如我们选取的两个 benchmark^[3,15];(2)它与 Euclidean 距离相比有较低的计算复杂度。

在第一阶段实验中,我们从图像库的每一类图像中分别随机地选取 20 幅图像,即共 200 幅图像作为查询图像。然后,计算其查准率,即所检索的相关图像与检索出的前 N 幅图像的比,在实验中 N 取 12,以便与选取的 benchmark^[3,15]和其它文献^[8,9]一致。实验结果如表 4 所示。

表 4 查准率(Precision)的实验结果

Image Categories	Benchmark-1	Benchmark-2	RLEBH
Cars	0.42	0.41	0.92
Flowers	0.41	0.60	0.67
Bike	0.62	0.33	0.60
Building	0.38	0.29	0.45
Animal	0.30	0.28	0.42
People	0.62	0.34	0.82
Mountains	0.28	0.31	0.43
Sky	0.39	0.53	0.47
Sunset	0.50	0.58	0.53
Textured	0.81	0.27	0.61
Average	0.47	0.39	0.59

表 5 查全率(recall)的实验结果

	Benchmark-1	RLEBH
顺时针旋转 10°	0.78	0.50
逆时针旋转 10°	0.74	0.52
缩小 20%	0.96	0.72
放大 20%	0.94	0.90
反色	0.02	1.00
水平翻转	0.92	0.98
垂直翻转	0.92	0.94
水平位移 10%	0.82	0.50
垂直位移 10%	0.86	0.52
加高斯噪声	0.02	0.90
加椒盐噪声	0.02	0.42
加乘性噪声	0.08	0.60
平均查全率	0.59	0.71
ANMMR	0.1	0.08

从上述实验结果可知:(1)本文提出的 RLEBH 相对于其它几种算法,其平均查准率最高。(2)RLEBH 算法的性能对大多数图像类而言都优于 benchmark-1,只有自行车和纹理两类图像的查准率稍次于 benchmark-1。这是因为 benchmark-1 的算法对图像的纹理特征有较好的辨识力,这得益于它从 DCT 域中提取均值和方差的过程中,获得了较多的图像纹理信息。(3)由于 benchmark-2 是基于色彩的内容描述符,RLEBH 算法的性能在色彩主导的图像类(如日落和天空)稍逊色于 benchmark-2,因为,在这两类图像中,彩色特征对图像内容的反映起到主导的作用。然而,对其它类的图像,RLEBH 算法的性能都高于 benchmark-2。

第二阶段的实验主要是验证本文提出的内容描述符在图像检索应用中的稳健性。从图像库中随机选取各种类型的图像 50 幅,并把每一幅图像都进行如下处理:顺时针旋转 10 度,逆时针旋转 10 度,缩小 20%,放大 20%,反色,水平翻转,垂直翻转,水平位移 10%,垂直位移 10%,加入高斯白噪声(均值为 0,方差为 0.04),加入椒盐噪声(方差为 0.04),加入乘性噪声(方差为 0.04),如图 2 所示。经上述处理后的图像再放回图像库中,用原图像作为查询图像在图像库中进行图像检索,然后计算查全率和文献[16]中定义的 ANMRR(average of the normalized modified retrieval rank)。

第二阶段的实验结果如表 5 所示。实验结果证明本文提出的内容描述符比 benchmark-1 在图像检索的平均查全率方面有更强的稳健性,在 ANMRR 方面也有更好的性能。这是因为:(1)块边缘模式是基于宏块的,从压缩域中提取,因此任何局部小区域的图像损伤或干扰,对单个像素值会带来较大

的影响,而对于整个宏块而言,这种影响则很小或可以忽略,因一个宏块包含 64 个像素;(2)当在一个宏块内的像素被损伤或干扰时,由于 DCT 变换的性质使其可以进行相互补偿,并且一个宏的块边缘模式是由 4 个 16 个像素组成的子块决定的,也就是说,个别像素值的变化,不会导致提取的块边缘模式发生改变。

结束语 本文提出一种有效的 JPEG 图像和 MPEG 视

频的内容描述符。与现行的算法比较,我们提出的算法具有如下优势:(1)对于传输、处理、编辑和存储过程中常常出现的图像方位变化和噪声干扰具有较好的稳健性;(2)RLEBH 是基于压缩域的算法,因此适合处理大量的压缩图像或视频;(3)极低的运算复杂度,便于实时实现和快速图像或视频内容管理应用。大量的实验证明,本文提出的算法与目前文献报导的类似算法相比,在平均查准率和查全率两方面都有所提高。

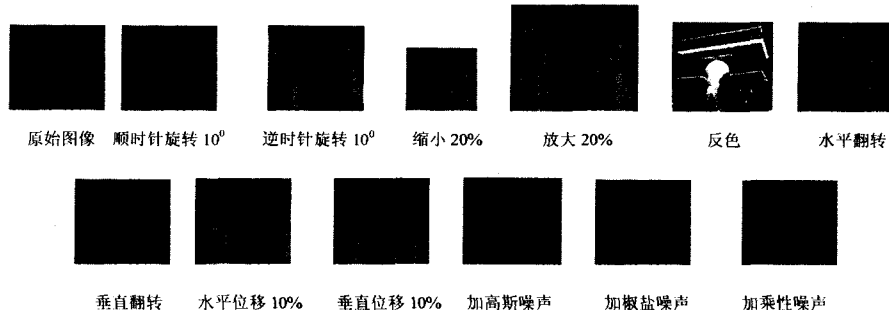


图 2 图像的各种变换实例

参考文献

[1] Chang S F. Compressed domain techniques for image/video indexing and manipulation// IEEE International Conference on Image Processing. 1995:314-317
 [2] Ngo C W, Pong T C. Exploiting image indexing techniques in DCT domain. Pattern Recognition, 2001, 34:1841-1851
 [3] Feng G Can, Jiang J M. JPEG compressed image retrieval via statistical features. Pattern Recognition, 2003, 36:977-985
 [4] Sim D G, Kim H K, Park R H. Fast texture description and retrieval of DCT-based compressed image. Electronic Letters, 2001, 37(1):18-19
 [5] Liu J H, H M Gu. Image retrieval in various domains. Computers & Graphics, 2003, 27:807-812
 [6] Zhong D, Defee I. DCT histogram optimization for image database retrieval. Pattern Recognition Letters, 2005
 [7] Soltane S, Kerkeni N, Angue J C. The use of two dimensional discrete cosine transform for an adaptive approach to image segmentation// Proceedings of the SPIE Image and Video Processing IV. 1996:242-251
 [8] Han J W, Guo L. A shape-based image retrieval method using salient edges. Signal processing: Image communication, 2003, 18:141-156
 [9] Banerjee M K, Kundu M. Edge based features for content based image retrieval. Pattern Recognition, 2003, 36: 2649-2661

[10] Kim D S, Lee S U. Image vector quantizer based on a classification in the DCT domain. // IEEE Trans. Commun, 1991, 39 (4): 549-556
 [11] Shen B, Sethi I K. Direct feature extraction from compressed images// Proc. SPIE: Storage and Retrieval for Still Image and Video Databases IV. 1996, 2670: 404-414
 [12] Chang H S, Kang K. A compressed domain scheme for classifying block edge patterns. IEEE Transactions on image processing, 2005, 14(2): 145-151
 [13] Li H, Liu G, Li Y. An effective approach to edge classification from DCT domain// Proc. IEEE Int. Conf. Image Processing. 2002:940-943
 [14] Jiang J. A low cost content adaptive and rate controllable near lossless image codec in DPCM domain. IEEE Trans. on Image Processing, 9(4): 543-554
 [15] Theo G, Arnold W M. PicToSeek: combining color and shape invariant features for image retrieval. IEEE Transactions on Image Processing, 2000, 9(1): 102-119
 [16] Bartsch H J. Handbook of mathematical formulas. Academic press, 1974
 [17] Popovici I, Withers W D. Custom-built moments for edge location. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4): 637-642
 [18] Shneier M, Mottaleb M A. Exploiting the JPEG compression scheme for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell, 18(8):849-853

(上接第 248 页)

地减小,增大 n 值取值范围,使产生的两个伪随机序列差值尽可能大,以保证水印提取端无法恢复水印信息。

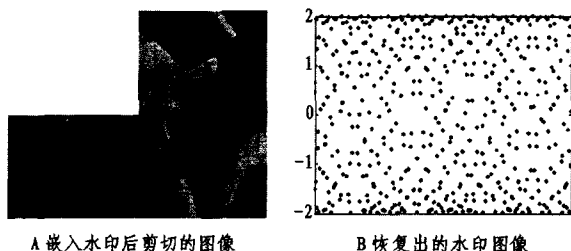


图 7 嵌入水印后剪切的图像与恢复出的水印图像

结束语 本文提出了基于分数阶微积分的分数阶伪随机序列数字水印算法。首先对正弦型信号进行分数阶微积分,再通过与其分数阶微分阶次相关的抽样,然后求得其差值伪随机序列 $v(n)$, $v(n)$ 对于分数阶微分的阶次 v 的初始值敏感性,保证了本分数阶伪随机数字水印的顽健性。

分数阶微积分在数字水印中应用的研究目前是一个新兴

的研究领域,本文做的工作只是在这一领域的一些初步的尝试,还有许多问题需要进一步研究,如:如何设计顽健性能更好的水印嵌入-提取算法。

参考文献

[1] 李忠源,郭全成,任亚萍. 图像中的信息隐含及水印技术[J]. 电子学报, 2000, 28(4): 61-63
 [2] Castleman K R. Digital Image Processing. Prentice-Hall, Inc, 1996
 [3] Oldham K B, Spanier J. The Fractional Calculus. New York and London: Academic Press, 1974
 [4] Yuan Xiao, Chen Xiangdong, LI Qiliang, et al. Differential operator and the construction of wavelet. Acta Electronica Sinica, 2002, 30(5): 769-773
 [5] Pu Yi-fei, Yuan Xiao, Liao Ke, et al. Five Numerical Algorithms of Fractional Calculus Applied in Modern Signal Analyzing and Processing [J]. Journal of Sichuan University (Engineering Science Edition), 2005, 37(5): 118-124
 [6] Trenevski K, Tomovski Z. On some fractional derivatives of functions of exponential type[J]. Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat., 2002, 13: 77-84
 [7] 胡广书. 数字信号处理: 理论与实现[M]. 北京: 清华大学出版社, 1997