

图像局部映射二进制串描述符

汤培楷*, 陈 玮, 麦毅成

(广东工业大学 自动化学院, 广州 510006)

(*通信作者电子邮箱 peikaitang@gmail.com)

摘 要: 为了提高图像局部特征算法的计算速率与匹配速度, 并保持其准确率, 提出了一种局部映射二进制串描述符算法。它通过映射的方法将图像局部区域转化成二进制串, 从而提高其计算效率与匹配速度, 并通过机器学习的方法寻找最佳映射矩阵, 保持其准确率。从实验结果表明, 只需 32 比特的二进制串就可以在准确率上媲美现有的局部描述符算法, 并在匹配速度上有较大的优势。

关键词: 二进制串; 局部特征; 描述符; 图像匹配; 特征提取

中图分类号: TP391.41 **文献标志码:** A

Binary projection for image local descriptor

TANG Peikai*, CHEN Wei, MAI Yicheng

(College of Automation, Guangdong University of Technology, Guangzhou Guangdong 510006, China)

Abstract: In order to reduce the computational burden and maintain the recognition rate of the image local descriptor, a binary projection method for image local descriptor was proposed. The image patch was projected and transformed into a binary string for boosting the performance as well as speeding up the matching speed. The projection matrix was optimized by machine learning method to maintain its recognition rate and robustness. The experimental result indicates that only a 32-bit binary string is needed to perform as well as the state-of-art descriptors and it shows significantly faster matching speed.

Key words: binary string; local feature; descriptor; image matching; feature extraction

0 引言

图像局部描述算子已广泛地应用于机器视觉领域, 如: 图像检索、姿态识别、三维重构、目标跟踪等。其主要功能是: 描述静态图像中的某个局部区域, 当图像发生变化时, 仍保持其不变性与鲁棒性, 并能快速有效地与其他图像区域进行匹配。在过去的十年里, 研究者针对其不变性与快速性提出了许多方法^[1-9]。

在这些方法中, 最著名的是 Lowe 在 2004 年提出的, 用梯度方向直方图来描述尺度不变特征变换 (Scale Invariant Feature Transform, SIFT)^[9] 特征, 得到 128 维的特征向量。在过去的几年里, 研究者主要的研究内容是: 如何尽可能地降低算法的复杂度, 同时可以在准确率上与 SIFT 相媲美。其中, Speeded up Robust Feature (SURF)^[7] 算法就是一个很好的例子, 它的匹配准确率与前者不相上下, 但计算时间缩短到原来的 1/3。它通过 Haar 小波响应求和来构造描述符, 使特征向量减少到 64 维, 其中每个维度用 4 个字节的浮点型表示。但是, 如果在实时视频跟踪、大型 3D 重构、图像检索等应用中运用这类算法时, 其运算效率、匹配速度远远达不到最低要求。2010 年 Calonder 等^[5] 提出了 Binary Robust Independent Elementary Features (BRIEF) 方法, 通过局部区域内像素点值大小的比较, 用二进制串的形式描述特征区域。此方法使算法的计算速度有很大的提高, 但由于其过于简单, 当图像发生旋转等变化时, 无法正确匹配。

二进制串描述符的优点在于: 只需用 0 或 1 就能对特征

区域进行描述。它生成描述符的计算速度更快, 更有利于计算机的存储。而且, 可以通过计算特征向量之间的汉明距离 (Hamming distance) 来比较特征的相似度。在基于二进制串描述符的算法中, 有的只是对旧算法 (如 SIFT) 的特征向量直接进行二进制编码, 这种方法依然会导致很大的计算量。而 Oriented Fast and Rotated BRIEF (ORB)^[12] 算法克服了 BRIEF 算子的缺点, 它通过比较特征区域内每一对小碎片的灰度值来得到二进制串; 并且对各个碎片的位置选择进行了优化处理, 得到相关度最低的碎片位置。但 ORB 采用穷举法来优化选择碎片的位置, 而且它对碎片大小的依赖性很高, 有较大的局限性。

本文提出一种局部区域映射的方法, 将特征区域图像的灰度值线性映射转化为二进制串, 并通过监督型机器学习的方法, 寻找线性映射矩阵的最优值, 生成只有 32 比特的二进制串, 在存储空间与匹配速度上有较大优势。

1 算法框架

对图像局部特征的提取, 主要分为两个步骤: 局部特征检测与局部特征描述。本文选用了最近几年比较流行的 FAST^[10] 方法作为基础, 来检测局部特征点。在局部特征描述阶段, 本文提出了一种基于二进制串的灰度值映射方法来描述特征区域, 在保证识别率的同时, 实现算法的快速性。

1.1 局部特征检测

FAST 特征点检测方法的计算速度高, 被广泛地应用于图像局部特征点检测。它通过比较中心点与周围区域的灰度值

收稿日期: 2012-10-09; 修回日期: 2012-11-29。

作者简介: 汤培楷 (1988 -), 男, 广东汕头人, 硕士研究生, 主要研究方向: 机器视觉; 陈玮 (1963 -), 女, 浙江温州人, 教授, 博士, 主要研究方向: 智能控制、离散事件动态系统; 麦毅成 (1989 -), 男, 广东佛山, 硕士研究生, 主要研究方向: 机器视觉。

来检测特征点。但 FAST 方法并不会生成多尺度特征,也无法确定特征点的主方向。本文把 FAST 应用到尺度金字塔,寻找多尺度特征点,计算其灰度值强度中心(Intensity centroid)^[11],并通过计算强度中心与特征点中心的偏移量,来确定主方向^[2]。其方法如下:

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \quad (1)$$

其中的 I 表示 x, y 坐标上图片的灰度值。由式(1)中的 m_{pq} , 可以找到强度中心点:

$$C = (m_{10}/m_{00}, m_{01}/m_{00}) \quad (2)$$

设特征中心点为 O , 则可以构造向量 OC 。那么其主方向 θ 可以表示为:

$$\theta = \arctan(m_{01}/m_{10}) \quad (3)$$

然后,把特征检测阶段得到的特征点位置、尺度、主方向的信息传递到特征描述阶段。

1.2 局部特征描述

在局部特征描述阶段,本文提出了一种线性映射的方法,将映射后的结果与阈值相比较,转换为二进制串来描述该区域的特征。

首先,将特征区域(32×32 的矩阵)的灰度值转换为一个列向量 x , 然后将其映射为多个数值,并与阈值相加,如下式所示:

$$\forall i \in 1, 2, \dots, N, p_i(x) = w_i^T x + t_i \quad (4)$$

其中: N 为二进制串的位数, w_i 表示第 i 个映射向量, t_i 为对应的阈值。最后,将其转化为二进制串:

$$b_i = \begin{cases} 1, & p_i(x) > 0 \\ 0, & p_i(x) \leq 0 \end{cases} \quad (5)$$

$$F = \sum_{1 \leq i \leq N} 2^{i-1} b_i \quad (6)$$

其中: F 为所求的二进制串特征描述符, b_i 为二进制串中每一位的值。算法的基本框架如图1所示。

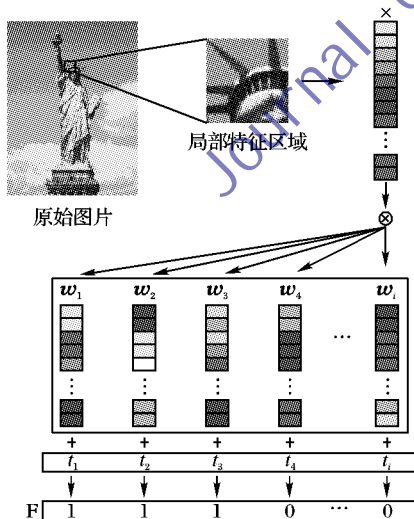


图1 映射二进制串描述符基本框架

2 算法学习训练方法

在构造二进制串映射描述符的过程中,最重要的步骤在于:如何找到合适的映射向量与阈值 $\{w_i, t_i\}$ 。在此基础上,选取前 N 组 $\{w_i, t_i\}$ 构成映射矩阵,优化描述符的匹配效率。本节将重点介绍描述符算法的学习优化方法。

2.1 求解映射向量与阈值

基本原则是,本文的方法是要使描述符在描述相同物体的局部特征时,其汉明距离必须最小化;而在描述不同物体的局部特征时,其汉明距离要最大化。从而得到一系列具有正交性质的线性映射向量 w_i 和与之相对应的阈值 t_i 。为了得到这个结果,同时采用了两个训练图库 A 与 B 来求解 $\{w_i, t_i\}$ 的最优值。其中 A 库里存放的都是相同物体的图像局部区域,而 B 中存放的都是不同物体的图像局部区域。

然而,直接对图像局部区域进行映射计算,会增加算法的计算复杂度。因此,本文设定所有映射向量 w_i 都是线性的,而且两两正交。为了进一步降低计算负荷,还希望 w_i 是一个稀疏向量, w_i 中的大多数元素都应该是0。

根据上述的算法优化思想,希望映射向量 w_i 在 A 图库中对相同物体的局部区域描述结果应该相同,即 $\text{sign}(p_i(x))$ 与 $\text{sign}(p_i(x'))$ 的值相等,在 B 图库则不相等,从而得到如式(7):

$$\min \sum_{i=1,2,\dots,N} \left[\sum_{(x,x') \in B} \text{sign}(w_i^T x + t_i) \text{sign}(w_i^T x' + t_i) - \sum_{(x,x') \in A} \text{sign}(w_i^T x + t_i) \text{sign}(w_i^T x' + t_i) + \alpha \|w_i\|_1 \right] \quad (7)$$

其中 w_i 是单位向量,而且满足线性正交性质,即

$$w_i^T \times w_j = 0; i \neq j \quad (8)$$

式(7)中的 $\alpha \|w_i\|_1$ 表示 w_i 的1-范数,可以通过调整 α 来改变 w_i 中元素的稀疏程度。

但是,式(7)中的 sign 函数是一个非线性不可微分函数,如果要通过式(7)来求解 $\{w_i, t_i\}$ 将会变得十分困难,较可行的办法是去掉其中的 sign 函数。由式(7)可知, w_i 和 t_i 是独立的变量,可以通过先求 w_i , 再求 t_i 的方法来求解 $\{w_i, t_i\}$ 。于是,可以先把 t_i 从式(7)中去掉,把式(7)近似表示为:

$$\min \sum_{i=1,2,\dots,N} \left[\sum_{(x,x') \in B} (w_i^T (x - x'))^2 - \sum_{(x,x') \in A} (w_i^T (x - x'))^2 + \alpha \|w_i\|_1 \right] \quad (9)$$

因为 w_i 是一个高维向量,而 t_i 只是一个数值。一旦得到 w_i 的求解, t_i 的求解将变得非常简单,只需要将求得的 w_i 代入式(7)中即可求解。

本文采用了随机梯度下降法(Stochastic Gradient Descent),即SGD来求解式(9)。SGD方法主要用于使目标函数最小化,其目标函数的类型为多个可微分函数的求和,符合式(9)的求解要求。然而,SGD方法具有一定的随机性,求解过程可能会陷入局部极小值。因此,只有合理设置迭代初始值 w_i^0 才能得到理想结果。本文参考了文献[3]的方法设置迭代初始值,加入了 w_i 单位正交化的限制条件,选择合适的训练图库。采用了由文献[12]提供的三个公用图像数据库:Liberty, Notre Dame 和 Yosemite,每个数据库都包含了超过40万张局部区域图片。所有图片都被标准化为 64×64 的灰度图,包含了尺度、主方向的相应信息。这些图片都采集于以关键点为中心的特征区域。数据库涵盖了图像光照、视点、仿射等多种变化,其中50%属于相匹配的区域,50%属于不匹配的区域。图2为Notre Dame数据库中的一张样例。以Notre Dame中前1万张局部图片作为训练样本,将其降采样为 32×32 ,通过Matlab求得一系列映射向量 w_i 。

2.2 选定二进制串的位数

通过对Notre Dame数据库前1万个特征区域进行学习训

练,得到了一系列 $\{w_i, t_i\}$, 按照其对应的目标函数值从小到大排列,选取前 i 组 $\{w_i, t_i\}$ 作为映射矩阵用于测试,从而选取最佳的二进制串位数。选用的测试图像库为 Yosemite 中的前 1 万个区域图片,设定合适的汉明距离阈值 Γ 。如果两个特征的汉明距离小于选定的阈值 Γ , 则认为两个二进制串特征是匹配的,反之则不匹配。根据结果记录不同位数(维度)的二进制串所对应的正确率,结果如图 3 所示,其正确率的计算方法如式(10)。

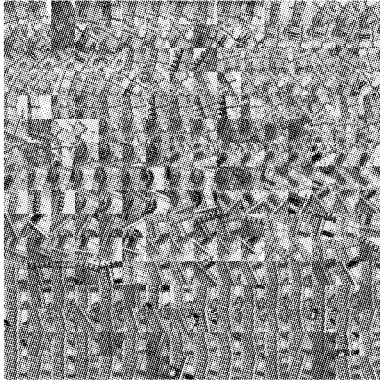


图2 Notre Dame 数据库中的一张样例

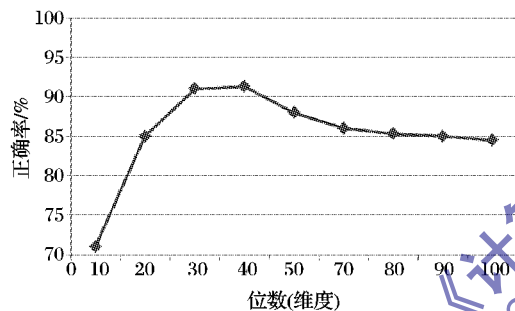


图3 选取前 N 组 $\{w_i, t_i\}$ 并测试其正确率

由图 3 可知,在选取前 30 到 40 组 $\{w_i, t_i\}$ 时,可以得到最高的正确率。随着维度的增大,正确率逐渐降低。这是因为选取的前 i 组 $\{w_i, t_i\}$ 是通过其对应目标函数值从小到大排列的。排列在越后面的 $\{w_i, t_i\}$ 所对应的目标函数值越大,会恶化描述符的整体性能。因此本文选用前 32 个 $\{w_i, t_i\}$ 作为映射矩阵。也就是说,只需要用 32 比特的二进制串就可以描述整个局部特征区域。

2.3 选定高斯模糊参数

在对局部区域进行映射计算之前,需要对该区域进行一次高斯模糊处理,从而减少某个噪声对整体图像的影响,增加特征描述的稳定性与重复性,有效地提高匹配准确率。本文参照了文献[5]中的方法,根据实验测试,选取合适的高斯模糊标准差 σ 。测试对象选用的是中介绍的图像库^[14],选取了其中一组 wall 图像对算法进行评估,找到最佳的标准差 σ 。Wall 组中有 6 张图片,其视点角度从 1 到 6 逐渐增大,选用第一张图片与其他 4 张图片进行测试,测试结果如图 4 所示,根据测试结果,选定 $\sigma = 2.0$ 作为高斯模糊滤波器的标准差。

3 性能测试比较

将使用训练后的映射描述符算法与当前比较流行的 SIFT、SURF、ORB 算法进行两个方面的比较:准确率与快速性。本文基于之前算法框架与机器学习训练结果,运用 OpenCV 库编写 C/C++ 程序,与 OpenCV 库中已存在的 SIFT、

SURF、ORB 算法进行比较。SIFT 生成的是 128 维的浮点型(float)特征向量;SURF 生成 64 维浮点型特征向量;而 ORB 作为二进制串描述符,生成 256 比特二进制串特征;本文的映射法描述符只生成 32 比特的二进制串特征。SIFT 在特征精确度上有很大优势,但这也降低其计算效率与匹配速度。

使用的计算机平台为: Intel Core i5 3.3 GHz CPU;4.0 GB RAM。

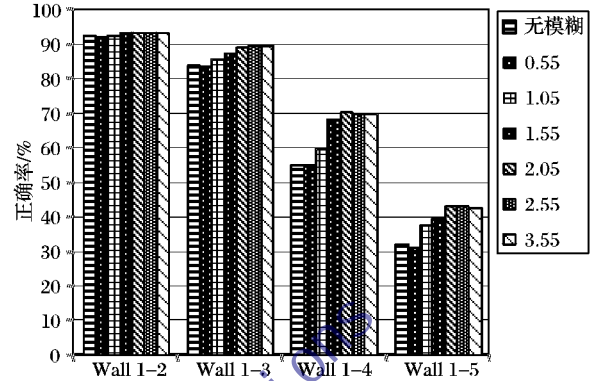


图4 选取不同 σ 值,测试 Wall1-2 到 Wall1-5 所得到的正确率

3.1 准确率比较

选用 Liberty 图像库的前 1 万张区域图像进行测试,采用了中推荐的方法:通过改变汉明距离(或欧拉距离)的阈值 Γ 来改变特征匹配的采纳度,即当两个特征的汉明距离(或欧拉距离)小于 Γ , 则认为两个特征是匹配的。因此, Γ 的取值越大,特征就越容易被匹配。每改变一次 Γ , 就可以得到图 5 中的一个点,测试结果如图 5 所示。

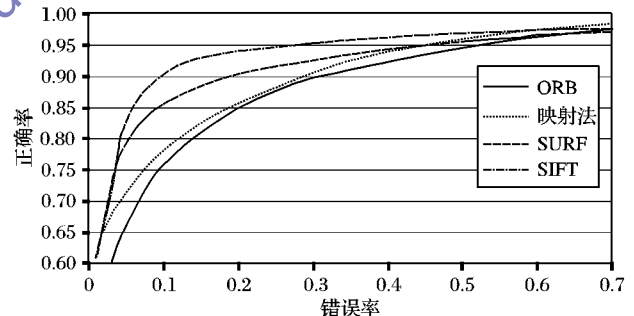


图5 测试 Liberty 图像库得到的错误与正确率对应曲线

图 5 中的正确率和错误率的计算方法如下:

$$\text{正确率} = \frac{\text{相同局部区域匹配个数}}{\text{相同局部区域的总数}} \quad (10)$$

$$\text{错误率} = \frac{\text{错误匹配个数}}{\text{匹配总数}} \quad (11)$$

由图 5 可以看出,本文的映射法描述符在准确率上低于 SIFT,但略高于 ORB 算法。这也是由于二进制串描述符为了减少特征存储空间,提高计算速度与匹配速度,需要以降低一定的准确率为代价。

3.2 计算速率比较

在一幅 1024×768 的图像上分别测试以上所说的四种描述符算法,每个算法都可以提取大约 1500 个特征点,经过 100 次重复测试,计算 SIFT、SURF、ORB 与映射法在计算每个局部特征时所需要的平均时间,结果如表 1 所示。

接着,测试各个算法的特征匹配速度。SIFT 与 SURF 采用的是欧氏距离法;ORB 与映射法采用的是汉明距离法,分别对 1 万对特征进行匹配运算,经过 10 次重复测试,计算每

次匹配所需的时间,结果如表1所示。

由表1可知,二进制串描述符在计算时间与匹配时间上,有巨大的提升。描述符计算效率最高的是ORB,略高于映射法,只需0.018 ms。匹配速度最快的是本文提出的映射法,其速度是SIFT的100倍,这也是本描述符算法最大的优点。

表1 描述符计算时间与描述符匹配时间比较

描述符算法	类型	计算时间/ms	匹配时间/ms
SIFT	128 维浮点型	0.501	0.412
SURF	64 维浮点型	0.161	0.201
ORB	256 比特串	0.018	0.008
映射法	32 比特串	0.034	0.004

4 结语

针对SIFT、SURF等传统描述符算法计算速度慢、匹配时间长的缺点,本文提出了一种基于映射法的二进制串描述符,它在计算速度与匹配速度上有较大的优势,匹配速度达到SIFT算法的100倍。而且此算法生成的局部特征只有32比特,在存储空间与匹配速度上优于同类的ORB描述符,可以在需要大量匹配运算的领域得到有效应用,如大规模图像检索、实时视频识别等。

参考文献:

- [1] 陈方,蒋云良,许允喜.改进的CenSurE特征和基于相加图像梯度的快速描述符[J].计算机应用,2011,31(7):1818-1821.
- [2] RUBLEE E, GARAGE W, PARK M, *et al.* ORB: an efficient alternative to SIFT or SURF [C]// IEEE International Conference on Computer Vision. Barcelona: IEEE, 2011: 2564-2571.
- [3] BROWN M, GANG H, WINDER S. Discriminative learning of local image descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(1): 43-57.
- [4] AMBAI M, YOSHIDA Y. CARD: Compact and real-time descrip-

tors [C]// IEEE International Conference on Computer Vision. Barcelona: IEEE, 2011: 97-104.

- [5] CALONDER M, LEPETIT V, STRECHA C, *et al.* BRIEF: binary robust independent elementary features [J]. ECCV'10: Proceedings of the 11th European Conference on Computer Vision. Berlin: Springer-Verlag, 2010: 778-792.
- [6] YU GUOSHEN, MOREL J-M. A fully affine invariant image comparison method [C]// ICASSP09: Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. Washington, DC: IEEE computer Society: IEEE, 2009: 1597-1600.
- [7] BAY H, TUYTELAARS T, van GOOL L. SURF: speeded up robust features [J]. Computer Vision - ECCV 2006. Berlin: Springer-Verlag, 2006: 404-417.
- [8] YAN K, SUKTHANKAR R. PCA-SIFT: a more distinctive representation for local image descriptors [C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE, 2004, 2: 506-513.
- [9] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [10] ROSTEN E, DRUMMOND T. Machine learning for high-speed corner detection [C]// Computer Vision - ECCV 2006. Berlin: Springer-Verlag, 2006: 430-443.
- [11] ROSIN P L. Measuring corner properties [J]. Computer Vision and Image Understanding, 1999, 73(2): 291-307.
- [12] Learning local image descriptors data [EB/OL]. [2012-11-28]. <http://www.cs.ubc.ca/~mbrown/patchdata/patchdata.html>.
- [13] MIKOLAJCZYK K, SCHMID C. A performance evaluation of local descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615-1630.
- [14] Robotics research group, university of Oxford. Affine covariant features [EB/OL]. [2012-11-28]. <http://www.robots.ox.ac.uk/~vgg/research/affine/>.

(上接第1073页)

参考文献:

- [1] 张小华,刘芳,焦李成.一种基于混沌序列的图像加密技术[J].中国图象图形学报,2003,8(4):374-378.
- [2] 李晓轩,王阿川.一种基于混沌排序的数字图像置乱算法[J].黑龙江科技信息,2009(27):54-54.
- [3] 赵学峰.基于面包师变换的数字图像置乱[J].西北师范大学学报:自然科学版,2003,39(2):26-29.
- [4] 顾勤龙,姚明.基于Logistic混沌序列的数字图像加密研究[J].计算机工程与应用,2003,23(3):114-116.
- [5] 郭建胜,金晨辉.对基于广义猫映射的一个图像加密系统的已知图像攻击[J].通信学报,2005,26(2):131-135.
- [6] 曹建秋,肖华荣.像素位置与像素值双重置换的混沌加密算法[J]. Computer Engineering and Applications, 2010, 46(28): 192-195.
- [7] WANG Y, WONG K W. A chaos-based image encryption algorithm with variable control parameters [J]. Chaos, Solitons & Fractals, 2009, 41(4): 1773-1783.
- [8] 孙鑫,易开祥,孙优贤.基于混沌系统的图像加密算法[J].计算机辅助设计与图形学学报,2002,14(2):136-139.
- [9] 张健,于晓洋,任洪娥.基于Cat映射和Lu混沌映射的图像加密

方案[J].电子器件,2007,30(1):155-157.

- [10] FU CONG, LIN BINBIN, MIAO YUSHENG, *et al.* A novel chaos-based bit-level permutation scheme for digital image encryption [J]. Optics Communications, 2011, 284(23): 5415-5423.
- [11] YE G D. Image scrambling encryption algorithm of pixel bit based on chaos map [J]. Pattern Recognition Letters, 2010, 31(5): 347-354.
- [12] ZHU Z L, ZHANG W, WONG K W, *et al.* A chaos-based symmetric image encryption scheme using a bit-level permutation [J]. Information Sciences: an International Journal, 2011, 181(6): 1171-1186.
- [13] 张雪峰,范九伦.基于位运算的数字图像隐藏技术[J].信息安全与通信保密,2007(5):149-150.
- [14] 许艳.基于位运算的BMP图像加密算法研究[J].湖南理工学院学报,2007,20(4):41-43.
- [15] 袁玲,康宝生.基于Logistic混沌序列和位换的图像置乱算法[J].计算机应用,2009,29(10):2681-2683.
- [16] 李涛,柳健.基于位平面与混沌系统的图像置乱方法[J].西南民族大学学报:自然科学版,2009,34(3):595-599.
- [17] 单佳佳,朱灿焰.基于位运算的图像加密技术的研究[J].电脑知识与技术,2007(5):804-805.