

基于 DCT 和 ONPP 的唇部特征提取

梁亚玲 杜明辉

(华南理工大学电子与信息学院 广州 510640)

摘 要 针对单视觉通道唇读系统中唇部特征的提取问题提出了基于 DCT+ONPP 的特征提取方法。相对于保持全局结构特性的 PCA 方法,ONPP 是一种既保持局部邻域几何特性又兼顾全局的线性降维技术。实验证明提出的方法优于 DCT+PCA 的特征提取方法。另外还对邻域点个数对系统性能的影响做了相应的研究,发现邻域点为 3 时该方法具有较好的识别效果。

关键词 唇部特征提取,正交邻域保持方法,唇读,主成分分析

中图法分类号 TP391

文献标识码 A

Lip Feature Extraction Based on DCT and ONPP

LIANG Ya-ling DU Ming-hui

(School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China)

Abstract For the visual feature extraction of visual only lipreading system, this paper proposed an DCT+ONPP method to extract the visual feature of lip, compared with PCA which preserving the global structure, ONPP is a method preserving the intrinsic geometry of the local neighborhoods, in the same time, the overlap of the neighborhoods can keep the global structure of the data. The experimental results show that the proposed method is better performance than PCA and DCT+PCA methods. About the influence of the neighbor numbers on the system, some research has been done on it, the experimental results show that 3 neighbors can get better performance.

Keywords Lip feature extraction, ONPP, Lip reading, PCA

1 引言

唇读是指通过观察说话者的口型变化,读出或者部分读出所说的内容^[1],它是人工智能研究的一个新方向,被广泛应用于噪声环境下提高语音识别率,如汽车内、机场、展馆及公共场所等。基于单视觉通道的语义识别系统在近几年迅速发展起来^[2-5],本文基于单视觉通道对唇读系统进行研究。一个典型的单视觉通道唇读系统可用图 1 表示。

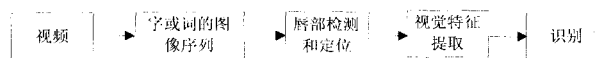


图 1 单视觉通道唇读系统示意图

如图 1 所示,单视觉通道唇读系统的实现过程为:输入表示唇部变化的视频,根据一定的规则将视频分割成具体的字或词的帧序列,经过唇部的检测和定位得到感兴趣区域,对感兴趣区域进行特征提取,利用提取的特征进行识别得到输入视频所代表的语义信息。由此可见,唇读系统要解决的主要问题为唇部检测和定位,唇部特征提取,训练和识别。由于特征提取对识别率影响非常大,因此本文针对唇部的特征提取进行研究。

唇部特征提取的方法大致可归为 3 类:(1)基于像素的方法。该类方法通过对图像的感兴趣区域即唇部图像进行某种

变换如 PCA, DCT, DWT 等,直接将变换后的相关参数作为唇部的特征;(2)基于轮廓的方法,该类方法主要考虑唇部的几何特征,如通过可变模板,主动形状模型或者通过曲线拟和的方法来获取唇部的几何特征;(3)基于像素和轮廓相结合的方法。该类方法将基于像素的方法和基于轮廓的方法获取的特征进行组合作为最终提取的特征。实际处理过程中,由于唇部几何特征受很多因素(人种、镜头、光线等)的影响,很难提取出较为理想的轮廓特征。因此目前广泛采用的是基于像素的方法,相关的研究^[6,7]表明基于像素的特征提取方法具有更高的识别率。本文的方法属于基于像素的方法。

基于像素的特征提取方法是对整个唇部区域(感兴趣区域)进行某种变换,得到特征的方法。一般来说该类方法获取的特征向量维数很高,通常存在大量冗余。文献[8]通过对感兴趣区域进行 DCT 变换,选取具有较高能量的 DCT 系数作为特征,由于该方法获取的特征仍然存在大量的冗余信息,文献[2,3]在文献[8]的基础上提出 DCT+PCA 的两级特征提取方法。该方法先对感兴趣区域图像进行 DCT 变换,然后用 PCA 对选取的 DCT 系数进行二次降维,提高了识别率。文献[4,5]在文献[2,3]的基础上提出了 DCT+LDA 方法。该方法将唇部口型分为 10 类,对能量较大的 DCT 系数根据 LDA 的方法选取最能区分唇部口型的特征,相对于 DCT 和

到稿日期:2010-06-28 返修日期:2010-11-01 本文受国家自然科学基金(U0735004)资助。

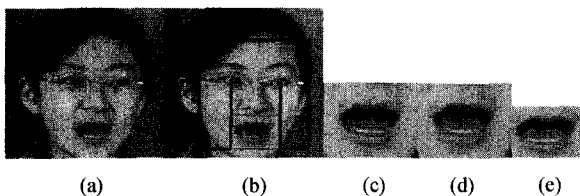
梁亚玲(1977—),女,博士生,讲师,主要研究方向为图像处理、图像编码,E-mail:ylliang@scut.edu.cn;杜明辉(1964—),男,博士,博士生导师,主要研究方向为信号处理、模式识别。

DCT+PCA 来说,该方法提高了唇读的识别率。文献[4,5]所描述的方法要求先对唇部口型分类,然后再用 DCT 及 LDA 提取特征。提取特征的目的并非为了实现语义的最佳分类,而是实现对不同口型的最佳分类,因此该方法受口形分类的影响较大。ONPP 是近几年发展起来的流行降维方法的一种,相对于 PCA 保持全局结构的特性,ONPP 具有保持数据局部几何特征和兼顾全局结构的特点。基于此,本文提出基于 DCT 和 ONPP 的唇部特征提取方法。

本文第 2 节介绍了唇部的检测和定位;第 3 节介绍了特征提取方法相关的理论;第 4 节介绍了本文方法;第 5 节给出实验,结果及分析;最后得出结论。

2 唇部的检测和定位

唇读系统中的感兴趣区域主要指唇部。一般指包含唇部的一个矩形区域,或者指包含唇部的下半部分脸。本文的感兴趣区域指包含唇部的矩形图像。唇部检测和定位的目的就是确定出感兴趣区域。一般唇部检测和定位是通过分析唇部色度与肤色的差别来确定感兴趣区域。这些方法一般受人种、光线、镜头的远近影响较大。获取的唇部感兴趣区域不能真实地反映唇部相关的特征。本文采用文献[9]的方法,该方法利用唇部的宽度与双眼之间距离相当的特点,首先利用 OpenCV 检测出人脸,确定人的双眼,然后根据双眼的距离以及嘴巴与眼睛之间的几何关系,确定出感兴趣区域的左右边界,假设人眼的连线与嘴巴是平行的,用双眼之间连线的角度将唇部调整至水平。再用文献[10]中的 LAB 空间 a 分量唇部提取方法对唇部的上下边界进行精确定位。通过归一化得到最终的 ROI 区域。该方法利用人脸中眼镜与嘴巴的关系来对 ROI 进行粗定位,对于特定人,该方法提取的感兴趣区域不受镜头远近、头部姿势等的影响,对光线变换也有较好的鲁棒性。图 2 给出具体实现过程的示意图。



(a) 输入图像, (b) 人脸和眼睛检测, (c) 唇部的粗定位, (d) 唇部方向调整, (e) 唇部精确定位, 分割和归一化后的 ROI

图 2 唇部的检测与定位

3 特征提取相关理论介绍

广义的特征提取是指通过某种变换,将样本从高维的采样空间转换到低维的特征空间,特征选择是指从一组特征中挑选出一些最为有效的特征以达到降低特征空间维数的目的。一般地,特征提取和特征选择紧密联系,很难划清界限,因此本文所说的特征提取是指二者的结合。

3.1 DCT 变换

离散余弦变换 (Discrete Cosine Transform, DCT) 具有很强的“能量集中性”,即经过变换后,输入信号(声音或图像)的能量主要集中在离散余弦变换后的低频部分。具体对图像来说,经过二维 DCT 变换后,其能量主要集中在 DCT 系数矩阵的左上方。DCT 的该特性被广泛地应用于信号和图像的有

损压缩和编码。用 DCT 进行特征提取,就是利用其能量集中这一特性,首先对图像进行二维 DCT 变换,再通过 zigzag 扫描将较大的系数排在前面,较小的系数排在后面。选取排在前面的系数作为图像的特征。

3.2 主成分分析 (PCA)

主成分分析 (Principal Components Analysis, PCA) 是指在数据中找出最“主要”的元素和结构,去除噪音和冗余,揭示隐藏在复杂数据背后的简单结构。对数据进行 PCA,实际上就是找到一种使全局重构误差最小的映射。将数据从采样空间映射到特征空间。因此 PCA 保持输入数据的全局特性。

3.3 ONPP 简介

正交邻域保持投影法 (Orthogonal Neighborhood Preserving Projections, ONPP)^[11]是由 E. Kokiopoulou 等提出来的基于流形的线性降维方法,在人脸识别和文字识别中都有很好的应用。相比 PCA 方法保持全局结构的特性,ONPP 可以保持局部邻域固有的几何特性,通过邻域的重叠及相互作用又可体现数据的全局几何特性。

给定数据集 $X = [x_1, x_2, \dots, x_n] \in R^{m \times n}$, 降维后特征向量 Y 的大小为 $d \times n$ 。每个特征向量的维数为 d 。用 ONPP 进行降维,就是将原始的数据集经过转换矩阵 V_{ONPP} 投影到低维特征空间。即

$$Y = U_{ONPP}^T X, U_{ONPP} \in R^{m \times d} \quad (1)$$

ONPP 假设每个采样点以及它的 k 个最邻近的点在一个局部的线性流形上。因此每个数据样本 x_i 可以通过它的 k 个最近邻线性组合的方式重构,重构的误差用最小化目标函数式(2)来表示。

$$\epsilon(W) = \sum_i \|x_i - \sum_j W_{ij} x_j\|_2^2 \quad (2)$$

式(2)中权值 W_{ij} 表示用邻域 $\{x_j\}$ 重建数据样本 x_i 的线性权值系数。ONPP 假设降维空间中的每个采样数据都可以通过它的 K 个邻域重构,而且它们具有同输入空间一样的权值,其采用的目标函数见式(3)。

$$\begin{aligned} \min_Y \epsilon(Y) &= \min_Y \sum_i \|y_i - \sum_j W_{ij} y_j\|_2^2 \\ &= \min_{U \in R^{m \times d}} \sum_i \|U^T x_i - \sum_j W_{ij} U^T x_j\|_2^2 \\ &= \min_{U \in R^{m \times d}} \sum_i \|U^T X (I - W^T)\|_2^2 \\ &= \min_{U \in R^{m \times d}} \text{tr}(U^T X M X^T U) \\ &= \min_{U \in R^{m \times d}} \text{tr}(U^T \tilde{M} U) \end{aligned} \quad (3)$$

式(3)中 $M = (I - W^T)(I - W)$, $\tilde{M} = X M X^T$, 如果给 U 附加一个条件,令 U 的列是正交的,那么 U 的解就是上述最优问题的特征向量。

整个 ONPP 的算法可以总结为:构造一个 K-NN 图,找到其 K 个邻域点;计算最佳线性重构输入数据 x_i 的权值 w_{ij} ;通过计算 $\tilde{M}$ 最小的特征值对应的特征向量得到投影矩阵 U ,对于输入的,计算其投影向量 $y_i = U^T x_i$ 。由于权值 w_{ij} 对于方向、尺度和平移都具有不变性,它的这些特性最直接的效果是保持了每个邻域内在的几何特性。

4 基于 DCT 与 ONPP 的唇部特征提取方法

如 3.2 节中所述,PCA 保持的是输入数据的全局特性。对于整体相似度较高、变化相对较小的唇部图像来说,局部特性非常重要。因此本文针对保持数据局部特性提出了基于

DCT+ONPP 的特征提取方法。该方法概括为:对输入的唇部图像,首先进行二维 DCT 变换,对变换后的系数进行 zigzag 扫描,选取前 100 个系数,用 3.3 节中所描述的 ONPP 的方法将输入数据投影到低维的特征空间,生成低维特征向量。ONPP 在降维的同时既保持了邻域之间的相互关系(通过构造邻域图),又通过邻域间的相互重叠间接体现了输入数据的整体的特性。对于输入的 DCT 系数来说,ONPP 降维所保持的是唇部图像不同频率 DCT 系数之间的相互关系,而该关系反映的是唇部图像能量和频率以及它们之间的关系,即唇部图像中,牙齿、皮肤、嘴唇以及阴影等所占比例及相互关系。

5 实验及其结果

5.1 识别方法

本文采用的识别方法是动态时间规整(Dynamic Time Warping, DTW)^[12]。DTW 是一种将时间规整和距离测度计算结合起来的一种非线性归整技术,它利用动态时间伸缩算法有效地解决了语音识别中,说话速率不均匀造成的时间伸缩问题,广泛地应用于孤立词的语音识别中。本文研究基于人的孤立词的唇读,采用 DTW 进行建模。

DTW 算法实质上是一种模板匹配的算法,设参考模板特征序列为: $A=a_1, a_2, \dots, a_j, \dots, a_m$, 测试序列特征矢量序列为 $B=b_1, b_2, \dots, b_j, \dots, b_n$ 。测试序列参数共有 n 帧矢量,参考模板共有 m 帧矢量。 B 和 A 可被映射到一个 $n \times m$ 的平面或者格中,每个格点 (i, j) 对应 b_i 和 a_j 的一个排列。动态时间归整算法就是要寻找一个最佳的时间归整函数 C , 使被测模板的时间轴 I 非线性地映射到参考模板的时间轴 J , 并使总的累计失真量最小。

$$DTW(A, B) = \min_c \left[\sum_{n=1}^N d(a_i(n), b_j(n)) \right] \quad (4)$$

式中, $d(a_i(n), b_j(n))$ 表示第 n 个匹配点对(参考模板的第 i (n) 个特征矢量和被测模板的第 j (n) 个特征矢量)之间的距离测度, $DTW(A, B)$ 则是在最优时间规整情况下被测模板和参考模板的距离。与测试模板距离最小的参考模板所对应的类即为识别结果。

5.2 实验

本文采用的数据库是 HIT BiCAVDatabase, 该数据库图像的采集速率为每秒 25 帧, 分辨率为 256×256 , 由 200 个句子组成, 约合 1000 个字, 每个句子录制 3 遍。在汉语中, 相同的音标会有不同的音调, 由于通过唇部口形不能区分不同音调, 因此本文将相同音标不同音调的视为同一个类。本文从该库的 1000 个字中选取了 96 个具有不同音标的字, 共 288 个样本进行实验。选取 DCT+PCA 作为对比算法进行了仿真实验。实验步骤如下:

(1) 采用第 2 节所述的唇部检测和定位方法对每幅图像进行预处理, 得到一个大小为 48×64 的唇部感兴趣区域图像。

(2) 对每幅唇部感兴趣区域图像进行二维 DCT 变换, 用 zigzag 扫描生成 1×3072 的向量, 选取前 100 个系数用 ONPP 的方法进行投影降维。

(3) 选取每个字第一个样本的特征作为 DTW 的模板, 另外的两个作为测试样本, 根据欧式距离得到其识别的结果。并统计其识别率。由于 ONPP 算法中邻域的个数会对识别率有一定的影响, 因此本文选取邻域个数分别为 3, 4, 5, 6 进

行了实验, 具体实验结果如表 1 所列。

表 1 实验结果

特征 维数	DCT+ PCA	DCT+ONPP (3)	DCT+ONPP (4)	DCT+ONPP (5)	DCT+ONPP (6)
40	62.50%	68.75%	67.71%	65.63%	66.67%
50	63.54%	69.79%	69.79%	70.83%	69.79%
60	62.50%	68.73%	65.63%	66.67%	65.63%
70	62.50%	69.79%	68.75%	65.63%	67.71%
80	62.50%	77.08%	76.04%	71.88%	70.83%
90	62.50%	75.00%	73.96%	70.83%	66.67%

表 1 中的特征维数指的是经过 PCA 降维或者是通过 ONPP 降维后的最终特征矢量的维数。对于 DCT+ONPP 方法, 括号中的数字表示邻域的个数。从表 1 可以看出, 本文方法在不同邻域个数的情况下, 识别率都要好于基于 DCT+PCA 的方法, 在邻域个数为 6 个时, 识别率平均提高了 5.21%。在邻域点数为 3 个时, 识别率平均提高了 9%。尤其是在邻域点为 3 个, 特征维数为 80 时, 识别率提高了 15%, 达到 77.08% 的识别率。这说明本文方法提出的唇部特征提取方法好于 DCT+PCA 特征提取方法。从侧面说明在唇部图像的特征提取过程中, 保持输入数据的局部特征对于识别具有非常重要的作用。对于不同的邻域个数, 识别率也有较大的差别, 从实验结果来看, 当邻域点的个数为 3 时其识别率最好。

由于文献[4,5]的识别率好于 DCT+PCA 的特征提取方法, 且为较新的方法, 因此有必要将本文的方法与文献[4,5]进行比较。由于文献[4,5]中未给出唇部口型分类的相关方法和原则, 也未给出不同口型的示意图, 因此本文无法对其进行仿真。这里只将本文方法的结果与文献[4,5]中列出的结果进行比较。

表 2 与文献[4,5]中最佳识别率的比较

	DCT+PCA	最佳识别率
文献[4,5]	65.2%(50)	68.4%(54)
本文	63.5%(50)	77.08%(80)

注: 表 2 中识别率后面括号内的数据表示达到最佳识别率时特征向量的维数。

从表 2 可以看出: 文献[4,5]对 DCT+PCA 方法进行仿真, 最佳识别率为 65.2%, 本文的仿真结果为 63.54%。由于唇读系统的识别率与 ROI 的检测与定位、特征提取、训练及识别方法都有关系, 因此结果不同属于正常。但本文与文献[2-5]仿真的结果均显示在特征维数为 50 维附近时可得到最高识别率。从这个角度来说, 直接将本文方法的实验结果与文献[4,5]的实验结果进行对比是具有可比性的。从最佳识别率来说, 本文方法取得了高于文献[4,5]约 9% 得识别率。这说明本文相对于文献[4,5]来说是一种有效的方法。另外本文方法在实现过程中不需要对唇部进行预先的分类, 因此在具体的实现过程中会方便很多。

结束语 基于流形的降维方法是最近几年迅速发展起来的一种降维方式, 其广泛应用于人脸识别等领域。本文应用正交的邻域保持投影来提取唇部的视觉特征, 相比 PCA 进行降维, ONPP 保持数据的局部几何特征, 对于非刚性唇部变化来说, 该特征是非常重要的。实验结果表明, 本文提出的 DCT+ONPP 的唇部特征提取方法在基于特定人的唇读识别率上有了较大程度的提高, 且在邻域数为 3 时达到较好的识别率。

参考文献

- [1] 姚鸿勋,高文,王瑞,等. 视觉语言-唇读综述[J]. 电子学报, 2001(2):239-246
- [2] Hong Xiao-peng, Yao Hong-xun, Wan Yu-qi, et al. A PCA based Visual Feature Extraction Method for Lip-Reading[C]// Proceeding of the 2006 international Conference on intelligent information Hiding and Multimedia Signal Processing(IIH-MSP'06)
- [3] 万玉奇,姚鸿勋,洪晓鹏. 唇读中基于像素的特征提取方法的研究[J]. 计算机工程与应用, 2007, 43(20):197-199, 221
- [4] 何俊,张华,刘继忠. 在 DCT 域进行 LDA 的唇读特征提取方法[J]. 计算机工程与应用, 2009, 45(32):150-152, 155
- [5] He Jun, Zhang Hua. Research on Visual Speech Feature Extraction [C]// 2009 international Conference on Computer Engineering and Technology. 2009:499-502
- [6] Atthaws M, et al. Extraction of visual features for lipreading [J]. IEEE Transaction on Pattern Analysis and Machine Intel-

- ligence, 2002, 24(2):198-213
- [7] Potamianos G, Neti C, Gravier G, et al. Recent advance in the automatic recognition of audio-visual speech[J]. Proceedings of the IEEE, 2003, 91(9):1306-1326
- [8] Heckmann M. DCT-based video features for audio-visual speech recognition[C]// Proc. Int Conf Spoken Lang Process Denver, USA, September 2002:1925-1928
- [9] Yao Wen-juan, Liang Ya-ling, Du Ming-hui. A Real-time Lip Localization and Tracking for Lip Reading[C]// 2010 the 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE 2010). Chengdu, China, August 2010:20-22
- [10] 梁亚玲,杜明辉. 基于 Lab 空间 a 分量的唇部提取方法[J]. 计算机工程, 2011, 37(3):19-21
- [11] Kokiopoulou E, Saad Y. Orthogonal Neighborhood Preserving Projections[C]// Proceeding of the Fifth IEEE International Conference on Data Mining(ICDM'05)
- [12] 易克初. 语音信号处理[M]. 北京:国防工业出版社, 2001

(上接第 260 页)

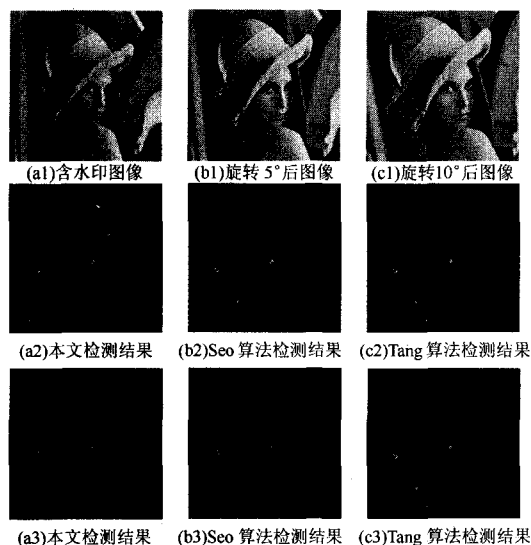


图5 不同算法抵抗旋转攻击比较

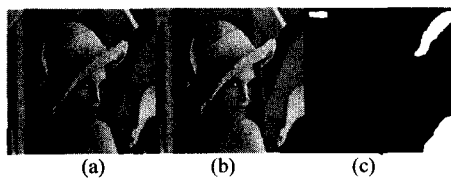


图6 (a)含水印 Lena 图像;(b)篡改后图像;(c)篡改检测结果

结束语 本文提出了一种基于图像特征的应用于图像版权保护和内容真伪鉴定的鲁棒性水印算法,即先使用 Hessian 仿射检测器获取宿主图的像特征区域,接着将版权保护信息嵌入到这些特征区域中,最后将脆弱水印嵌入到图像的非特征区域。对比其他水印算法,本文算法具有更好的鲁棒效果,并且可以通过脆弱水印信息对图像所遭受的篡改位置进行定位。由于图像特征区域具有仿射不变性,故该算法能抵抗一些常规的如旋转、缩放、裁剪等几何攻击。但是由于 Hessian 仿射检测过程计算量较大,使得整个算法的计算复

杂度较高。因此,在以后的实验中可以考虑将脆弱水印信息嵌入到图像的特征区域以增加算法精度。

参考文献

- [1] Lin C Y, Wu M, Bloom J A, et al. Rotation, scale and translation resilient water-marking for image[J]. IEEE Trans. Image Processing, 2001, 10(5):767-782
- [2] Jin C, Wang S H. Robust Digital Watermark Technique for Copyright Protection[J]. Information Engineering and Electronic Commerce, 2009, 3(10):237-240
- [3] Tang C W, Hang H M. A feature-based robust digital image watermarking scheme[J]. IEEE Trans. Signal Processing, 2003, 51(4):950-959
- [4] Seo J S, Yoo C D. Localized image watermarking based on feature points of scale-space representation[J]. Pattern Recognition, 2004, 37(7):1365-1375
- [5] Lee H Y, Kim H, Lee H K. Robust image watermarking using local invariant features[J]. Journal SPIE, Optical Engineering, 2006, 45(3)
- [6] Shu Z H, Liu G D. Image Watermarking Based on Wavelet-Based Contourlet Packet Transform with Best Tree[J]. IEEE Trans. Image Processing, 2009, 1(10):203-207
- [7] Mikolajczyk K, Schmid C. Scale and affine invariant interest point detectors[J]. Int. J. Computer Vision, 2004, 60(1):63-86
- [8] Mikolajczyk K, Tuytelaars T, Schmid C, et al. A comparison of affine region detectors[J]. Int. J. Computer Vision, 2005, 65(1/2):43-72
- [9] 王向阳,侯丽敏. 基于图像特征点的强鲁棒数字水印嵌入方案[J]. 自动化学报, 2008, 34(01):1-6
- [10] 刘红翼,王继军. 一种基于 LSB 的数字图像信息隐藏算法[J]. 计算机科学, 2008, 28(01):102-125
- [11] 张宪海,杨永田. 基于脆弱水印的图像认证算法研究[J]. 电子学报, 2007, 35(01):34-39
- [12] 王向阳,杨艺萍. 基于小波矩的抗几何攻击数字图像水印算法研究[J]. 中国图像图形学报, 2010, 05(01):15-19