

国家自然科学基金项目资助(61379143, 61472424, 61273143)

徐州市科技计划项目资助(XM13B119)

图像质量评价中的 特征提取方法与应用

李雷达 著

中国矿业大学出版社

China University of Mining and Technology Press

国家自然科学基金项目资助(61379143,61472424,61273143)
徐州市科技计划项目资助(XM13B119)

图像质量评价中的特征 提取方法与应用

李雷达 著

中国矿业大学出版社

内 容 提 要

随着各种各样便携式拍照设备的出现,尤其是配备了高分辨率摄像头的智能手机的广泛普及,数字图像在人们生产和生活中的应用越来越广泛,在现代信息社会中扮演着越来越重要的角色。本书结合著者在图像质量评价方向的研究,重点介绍数字图像的客观质量评价中常用的特征提取方法,并将这些特征提取方法用于设计图像质量评价模型。全书内容主要包括:绪论、图像质量评价经典方法、数字图像特征提取、全参考图像质量评价方法研究、无参考型 JPEG 图像块效应评价方法研究、无参考型图像模糊评价方法研究、总结与展望。

本书可供相关专业的研究人员借鉴、参考,也可供广大教师和学生使用。

图书在版编目(CIP)数据

图像质量评价中的特征提取方法与应用/李雷达著.

徐州:中国矿业大学出版社,2015.6

ISBN 978 - 7 - 5646 - 2711 - 9

I. ①图… II. ①李… III. ①数字图像处理—研究
IV. ①TN911.73

中国版本图书馆 CIP 数据核字(2015)第128851号

书 名 图像质量评价中的特征提取方法与应用

著 者 李雷达

责任编辑 何晓明

出版发行 中国矿业大学出版社有限责任公司

(江苏省徐州市解放南路 邮编 221008)

营销热线 (0516)83885307 83884995

出版服务 (0516)83885767 83884920

网 址 <http://www.cumtp.com> E-mail: cumtpvip@cumtp.com

印 刷 徐州中矿大印发科技有限公司

开 本 850×1168 1/32 印张 5.875 字数 155 千字

版次印次 2015 年 6 月第 1 版 2015 年 6 月第 1 次印刷

定 价 32.00 元

(图书出现印装质量问题,本社负责调换)

前 言

随着各种各样便携式拍照设备的出现,尤其是配备了高分辨率摄像头的智能手机的广泛普及,数字图像在人们生产和生活中的应用越来越广泛,在现代信息社会中扮演着越来越重要的角色。相应地,学术界针对数字图像的研究也越来越深入和广泛。数字图像在采集、处理、传输等各个环节中都可能引入不同类型的失真,导致图像质量的下降。均方误差(MSE)和峰值信噪比(PSNR)是目前最常用的两种图像质量评价方法,它们计算简单且具有明确的物理意义。但同时,它们的评价结果与人眼的主观感受存在一定的差异,因而不能很好地反映图像的客观质量。此外,它们评价图像质量时需要参考图像,这在现实环境中常常不能满足。针对这种情况,数字图像的客观质量评价方法在学术界和产业界引起了广泛的关注,已经成为图像处理领域的一个热门研究课题。

本书结合笔者在图像质量评价方向的研究,重点介绍数字图像的客观质量评价中常用的特征提取方法,并将这些特征提取方法用于设计图像质量评价模型。主要包括:

第1章:绪论。主要介绍图像质量评价的研究背景与意义、方法分类、现有图像质量数据库和衡量图像质量评价算法性能的主要标准等。

第2章:图像质量评价经典方法。针对全参考型、部分参考型和无参考型三类方法,分别介绍其主要评价过程,并简单介绍近年来出现的具有代表性的算法。

第3章:数字图像特征提取。针对图像质量评价,介绍几类特

征提取方法,包括图像中的边缘与梯度、局部不变特征、正交矩和稀疏表示等。

第4章:全参考图像质量评价方法研究。介绍基于离散正交矩和多尺度结构表达的两种全参考型图像质量评价算法,在标准图像质量数据库中进行性能测试,并与现有经典算法性能进行对比。

第5章:无参考型 JPEG 图像块效应评价方法研究。介绍基于 JPEG 图像中颜色漂移特性、基于 Tchebichef 核函数分析、基于 JPEG 网格强度与规律性的图像中块效应的无参考评价算法,在标准图像质量数据库中的 JPEG 图像上进行性能测试,并与现有经典算法性能进行对比。

第6章:无参考型图像模糊评价方法研究。介绍基于稀疏表示和离散正交矩的两种无参考型图像模糊评价算法,在标准图像质量数据库中的模糊图像上进行性能测试,并与现有经典算法性能进行对比。

第7章:总结与展望。对本书的主要内容进行总结,并对视觉质量评价今后的研究进行展望。

本书的出版得到了国家自然科学基金项目(61379143, 61472424, 61273143),徐州市科技计划项目(XM13B119),中央高校基本科研业务费专项资金(2015QNA66, 2014YC07, 2013RC10, 2013RC12),上海市信息安全综合管理技术研究重点实验室开放课题(AGK2012002)和中国博士后科学基金特别资助项目(201104586)等资助,在本书的撰写过程中参考并引用了国内外其他研究者的部分论文,在此一并表示感谢。

笔者在本书的撰写过程中竭尽全力,但由于水平有限,难免有错误之处,恳请读者批评指正。

著者

2014年10月

目 录

1 绪论	1
1.1 研究背景与意义	1
1.2 图像质量评价方法分类	4
1.3 图像质量评价数据库	6
1.4 图像质量评价算法的性能衡量标准	10
1.5 本书内容安排	11
2 图像质量评价经典方法	14
2.1 全参考型图像质量评价	14
2.1.1 SSIM	15
2.1.2 MS—SSIM	17
2.1.3 VIF	17
2.1.4 FSIM	18
2.1.5 GSM	19
2.1.6 GMSD	20
2.1.7 VSI	20
2.2 部分参考型图像质量评价	22
2.3 无参考型图像质量评价	24
2.3.1 块效应无参考评价	24
2.3.2 模糊无参考评价	28
2.3.3 通用型无参考质量评价	32
2.4 本章小结	34

3 数字图像特征提取	35
3.1 边缘	35
3.1.1 梯度	35
3.1.2 Roberts 算子	36
3.1.3 Prewitt 算子	36
3.1.4 Sobel 算子	37
3.1.5 LoG 算子	37
3.1.6 Canny 算子	39
3.2 局部不变特征	40
3.2.1 Harris 特征点	41
3.2.2 多尺度特征点	42
3.3 正交矩	49
3.3.1 连续正交矩	49
3.3.2 离散正交矩	54
3.4 高层次特征	58
3.4.1 稀疏表示	58
3.4.2 字典学习	59
3.5 本章小结	61
4 全参考图像质量评价方法研究	62
4.1 基于离散正交矩的全参考图像质量评价算法	62
4.1.1 Wee 算法与分析	62
4.1.2 改进的算法	64
4.1.3 实验结果与分析	66
4.1.4 小结	71
4.2 基于多尺度结构表示的全参考图像质量评价 算法	73
4.2.1 相关方法	73

4.2.2	多尺度高斯差 (DoG)	74
4.2.3	本节的算法	77
4.2.4	实验结果与分析	81
4.2.5	小结	86
4.3	本章小结	86
5	无参考型 JPEG 图像块效应评价方法研究	88
5.1	基于颜色漂移特性的 JPEG 图像块效应盲评价 算法	88
5.1.1	JPEG 图像中的颜色漂移与颜色丢失	89
5.1.2	基于颜色漂移特性的块效应评价	90
5.1.3	实验结果与分析	96
5.1.4	小结	100
5.2	基于 Tchebichef 矩的图像块效应盲评价算法	100
5.2.1	Tchebichef 核函数与块效应的关系	100
5.2.2	块效应评价	102
5.2.3	实验结果与分析	104
5.2.4	小结	109
5.3	基于网格强度与规律性的 JPEG 图像块效应盲 评价算法	110
5.3.1	块效应网格提取	111
5.3.2	网格强度描述	113
5.3.3	网格规律性描述	114
5.3.4	图像块效应分数	117
5.3.5	实验结果与分析	117
5.3.6	小结	125
5.4	本章小结	125

6 无参考型图像模糊评价方法研究	126
6.1 基于离散正交矩的图像模糊评价算法	127
6.1.1 Tchebichef 矩与图像模糊的关系	128
6.1.2 图像模糊评价	130
6.1.3 实验结果与分析	135
6.1.4 小结	143
6.2 基于稀疏表示的图像模糊评价算法	145
6.2.1 稀疏表示与模糊的关系	145
6.2.2 模糊评价算法	146
6.2.3 实验结果与分析	150
6.2.4 小结	160
6.3 本章小结	161
7 总结与展望	163
7.1 内容总结	163
7.2 研究展望	164
参考文献	166

1 绪 论

1.1 研究背景与意义

人类从外界感知的信息中有 60% 以上来自于视觉, 可见视觉信息在我们日常生活中的重要性。随着计算机技术、网络与多媒体技术的飞速发展, 数字图像在我们生活中的方方面面体现出越来越重要的作用, 包括医学、教育、气象学、军事、通信、影视等。尤其是近年来, 智能手机已经配备有分辨率相当高的摄像头, 加之社交网络的快速普及和流行, 数字图像的分享轻而易举。同时, 现实的需求也在默默地推动图像处理相关领域的发展, 针对数字图像的各种处理算法和技术不断涌现, 更丰富了数字图像作为一种重要信息载体的功能。可以说, 在现代社会生产和生活的每个方面, 我们都离不开数字图像和图像处理技术。

在现实环境中, 我们总是希望获取和应用高质量的数字图像, 而在图像的获取、处理、传输等一系列环节中, 都不可避免地引入各种不同类型的失真, 从而造成图像质量的下降。因此, 如何评价图像的质量一直以来都是一个重要的研究课题。在数字图像处理技术中, 应用最广泛的两种图像质量评价方法为均方误差 (Mean Squared Error, MSE) 和峰值信噪比 (Peak Signal to Noise Ratio, PSNR)。给定两个相同大小的待比较图像 $f_1(x, y)$ 和 $f_2(x, y)$, $x \in \{1, 2, 3, \dots, M\}$, $y \in \{1, 2, 3, \dots, N\}$, MSE 和 PSNR 的定义如下:

$$\text{MSE} = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N [f_1(x, y) - f_2(x, y)]^2 \quad (1-1)$$

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{\text{MSE}} \right) \quad (1-2)$$

式中, 255 表示数字图像灰度的最大值。两幅图像越接近, MSE 的值越小, 相应的 PSNR 的值就越大。

从二者的定义中可以看出, MSE 和 PSNR 在计算的时候主要是求两个图像中对应位置像素的灰度值差异。这种计算方法简单, 而且它们有较为明确的物理含义, 即衡量图像的灰度值差异。但是在实际应用中, 上述的图像质量描述方式存在以下问题^[1]:

(1) 人眼感觉图像差异的能力有限, 很小的像素差异人眼并不能觉察到, 因此并不影响人眼对图像质量的感受。在这种情况下 MSE/PSNR 值并不能很好地反映图像的主观质量。

(2) 人眼在观察图像的时候, 对图像中不同区域的关注程度不一样, 因此不同区域的失真对视觉质量的重要程度不一样, 而 MSE/PSNR 将图像中的不同像素/区域同等对待, 因此不能很好地反映人眼观察的结果。

(3) 对图像的操作并不一定都会造成质量的下降, 例如现存的各种图像增强算法在对图像操作后并不会造成图像质量的下降, 因此在这种情况下计算得到的 MSE/PSNR 值并不能很好地反映图像的实际质量。

(4) 由于视觉掩蔽特性, 对图像的修改在不同的区域会产生不同程度的视觉感受, MSE/PSNR 在计算的时候并没有考虑这些因素, 因此无法客观地反映像素变化对视觉质量的影响。

(5) MSE 和 PSNR 在计算的时候需要参考图像, 即全参考图像质量评价方法, 在实际中如果只有失真图像, 则无法使用 MSE/PSNR 评价图像质量。

出于以上原因,近年来学术界开始研究符合人眼视觉系统特性的图像质量评价方法,以求用更加有效的数学模型来描述图像质量,同时力图保持与人眼主观感觉的一致性。

图像质量评价技术在现实环境当中有很多用处,典型的应用包括^[2]:

(1) 图像处理算法性能比较。如今,针对一种特定的图像处理技术,例如图像压缩、去噪、超分辨重构、自适应缩放等,现有的文献中都提出了各种各样的算法。利用图像质量评价算法,能更为客观地对算法的性能进行评估,从而确定哪一种算法能获得更好的处理效果。

(2) 图像处理算法参数优化。在一种特定的图像处理算法中,往往存在多个参数,这些参数对算法的性能有着不同的影响。利用图像质量评价方法,能够系统地测试不同的参数变化对最终图像质量的影响,在此基础上可以进一步对各个参数进行优化,以获得最佳的算法性能。

(3) 网络环境下的服务质量(Quality of Service, QoS)控制。图像和视频在网络环境中传输时,会不可避免地造成一定程度的失真,因此需要对 QoS 进行控制。在这种情况下,由于往往没有原始的图像或视频可以作为参考,因此需要能够直接利用所观测到的图像和视频进行质量评价,获得图像的质量信息,进而作为反馈信息用于提高 QoS。

(4) 图像识别算法性能评估。在模式识别领域,有很多课题研究是针对图像信号进行的,如人脸识别、指纹识别、目标跟踪等。这些图像/视频信号在获取的时候,往往会伴随不同程度的失真,包括图像压缩、噪声、光照等的影响。利用质量评价算法,可以衡量图像在产生不同程度的失真时,图像质量与识别精度的关系。在此基础上,可以通过对图像质量的控制,来保证实际应用中所需要的识别精度。

(5) 医学图像压缩。不同于一般的图像,医学图像一般具有更高的分辨率,其存储也需要更大的空间,而且由于医学图像一般用于对病人病情的诊断,所以对医学图像的压缩和处理应以不影响对病情的诊断作为前提。利用图像质量评价算法,可以在保证图像质量的前提下,通过控制压缩算法的参数实现最大程度的压缩,节省存储空间。

1.2 图像质量评价方法分类

对图像质量的评价,可以分为主观评价方法和客观评价方法。人眼是图像的最终感知者和评价者,所以图像质量最可靠的评价方法是人直接观察。图像质量主观评价方法主要分为两种,即绝对评价和相对评价。绝对评价不需要与其他图像对比,凭人的感觉直接判断图像质量的好坏。这种评价方法的评价尺度见表 1-1。

表 1-1 图像质量绝对评价尺度

质量尺度		妨碍尺度	
5 分	看不出图像质量下降	5	非常好
4 分	能看出图像质量变化,但不妨碍观看	4	好
3 分	能看出图像质量损坏,对观看稍有妨碍	3	一般
2 分	对观看有妨碍	2	差
1 分	严重妨碍观看	1	非常差

相对评价需要与一组其他图像进行对比以判断图像质量,这种方法的评价尺度见表 1-2。

表 1-2 相对评价尺度和绝对评价尺度对照表

分数	相对测量尺度	绝对测量尺度
5 分	一群中最好的	非常好
4 分	好于该群中平均水平的	好
3 分	该群中的平均水平	一般
2 分	差于该群中平均水平的	差
1 分	该群中最差的	非常差

图像质量主观评价方法简单直接,但是在实际中容易受到环境、观察者的状态、知识背景等各种因素的影响,可重复性差。此外,如果有大量的图像需要处理,主观评价方法效率低,因此不适合于在线处理。在这种情况下,就需要自动化的客观评价算法。

图像质量客观评价方法通过建立数学模型对图像的质量进行定量评价,同时通过各种方式使预测得到的质量分数与人的主观感觉尽可能保持一致。现有的客观图像质量评价方法,根据对参考图像的依赖程度,可以分为^[1,3,4]全参考型(Full-reference, FR)、部分参考型(Reduced-reference, RR)和无参考型(No-reference, NR)。这三种方法的工作原理如图 1-1 所示。

全参考型方法在评价图像质量的时候需要用到测试图像和参考图像的全部信息。部分参考型方法在评价时只需要从原始图像中提取边信息,通过将测试图像中信息与上述边信息进行比较,即可确定图像质量。全参考方法和部分参考方法在评价时都需要参考图像的信息(全部或部分)进行对比,而在实际环境中这些信息可能没有或者很难获得,因此这两类方法的应用范围有限。相比之下,无参考型方法仅利用失真图像本身来计算质量分数,因此更符合实际环境的需要。当然,由于仅有失真图像可用,无参考图像质量评价方法的设计也更具挑战性。

无参考型图像质量评价方法还可以进一步分为针对特定失真

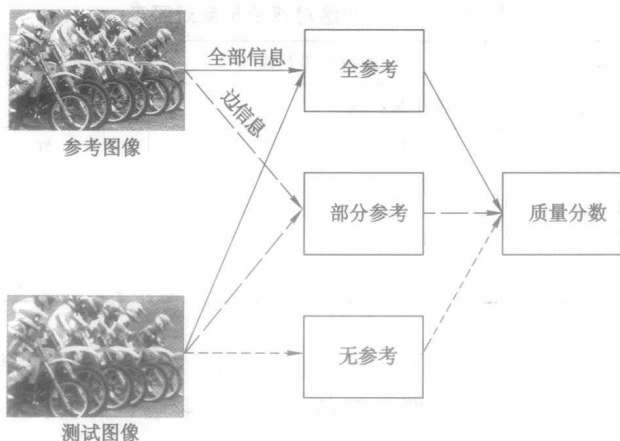


图 1-1 客观图像质量评价方法分类

的算法和通用型算法。前者主要针对图像中的某一种失真进行评价,如块效应、模糊、振铃效应等。通用型无参考方法在评价图像质量时,并不区分图像中存在的具体是何种失真,而是进行综合评价。这种通用型的无参考图像质量评价方法在实际中的用途更广泛,也是近年来研究的热点。

在过去的几年中,出现了大量的图像质量评价算法。这些算法采用不同的特征提取技术,从不同的角度挖掘图像质量评价中的关键信息,同时结合人眼视觉系统的特性,生成与主观评价具有较好一致性的图像质量分数。

1.3 图像质量评价数据库

为了有效地衡量图像质量评价算法的性能,需要将算法在标准的图像质量数据库中进行测试。目前,常用的 8 个图像质量数据库包含: LIVE^[6-8]、CSIQ^[9]、TID2008^[10]、TID2013^[11]、

MICT^[12]、IVC^[13]、A57^[14]和 WIQ^[15,16]。在这些数据库中,一般包含高质量的参考图像,对参考图像添加了不同类型和强度失真的图像,以及对失真图像的主观质量分数。这些主观质量分数主要通过人眼观察并求平均得到。通常,衡量主观质量的标准有平均意见分数(Mean Opinion Score, MOS)和差分平均意见分数(Differential Mean Opinion Score, DMOS)。对 MOS,分数越高表示图像的质量越好;对 DMOS,分数越低表示图像质量越好。下面针对各个数据库进行简单介绍:

(1) LIVE:由 University of Texas 的 Laboratory for Image and Video Engineering (LIVE)实验室开发。该数据库中包含 29 幅参考图像和 779 幅失真图像,所有图像均为彩色 BMP 格式,图像的分辨率从 634×438 到 768×512 大小不等。此数据库中共包含 5 种类型的失真,具体为高斯模糊(145 幅)、加性高斯白噪声(145 幅)、JPEG 压缩(169 幅)、JPEG2000(175 幅)和模拟 Rayleigh 衰减信道失真(145 幅)。对每幅图像,每一种失真都存在 5~6 种不同的强度。每幅图像都由 20~29 个人打分,然后将打分转换为 1~100 之间的分数,最后对每一幅图像用 DMOS 表示其主观质量分数。

(2) CSIQ:由 Oklahoma State University 的 Computational Perception and Image Quality Lab 开发。该数据库包含 30 幅参考图像和 866 幅失真图像,所有图像均为彩色 PNG 格式,图像的分辨率为 512×512 。此数据库中共含有 6 种类型的失真,包括 JPEG 压缩(150 幅)、JPEG2000 压缩(150 幅)、加性高斯白噪声(150 幅)、加性高斯粉色噪声(150 幅)、高斯模糊(150 幅)和整体对比度降低(116 幅);每一种失真有 4~5 种不同的强度。共有 35 个人参与图像主观质量的打分,最后用 DMOS 表示图像的主观质量分数。

(3) TID2008:由 Tampere University of Technology 开发。

该数据库包含 25 幅参考图像和 1 700 幅失真图像,所有图像均为彩色 BMP 格式,图像的分辨率为 384×512 。该数据库中共含有 17 种类型的失真,每幅图像的每种失真有 4 个不同的强度。这 17 种失真分别为:加性高斯噪声、色彩分量比亮度分量强度更大的加性噪声、空间相关噪声、掩膜噪声、高频噪声、脉冲噪声、量化噪声、高斯模糊、图像去噪、JPEG 压缩、JPEG2000 压缩、JPEG 传输误差、JPEG2000 传输误差、非偏心模式噪声、不同强度的局部块失真、亮度偏移和对比度改变。共有 838 个人参与了主观质量打分,用 MOS 表示图像的主观质量分数。

(4) TID2013:在 TID2008 的基础上开发而来,是目前图像数最多的数据库,包含 25 幅参考图像和 3 000 幅失真图像,所有图像均为彩色 BMP 格式,图像的分辨率为 384×512 。该数据库中共含有 24 种类型的失真,每幅图像的每种失真有 5 个不同的强度。除了包含 TID2008 中的 17 种失真之外,增加了 7 种新的失真类型,具体为:色彩饱和度改变、乘性高斯噪声、舒适噪声、含噪图像的有损压缩、带抖动的图像颜色量化、色差、稀疏采样与重构。共有来自 5 个不同国家的 971 个人参与了主观质量打分,用 MOS 表示图像的主观质量分数。

(5) MICT:由 University of Toyama 的 Media Information and Communication Technology Lab 开发。该数据库包含 14 幅参考图像和 168 幅失真图像,所有图像均为彩色 BMP 格式,图像的分辨率为 512×768 。该图像库只含有两种类型的失真,即 JPEG 压缩(84 幅)和 JPEG2000 压缩(84 幅)。共有 16 个人参与了图像主观质量的打分,用 MOS 表示图像的主观质量分数。

(6) IVC:由 Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN)开发。该数据库包含 10 幅参考图像和 185 幅失真图像,图像为彩色格式,分辨率为 512×512 。