

现代图像质量评价

Modern Image Quality Assessment

【加】王周 【美】阿兰·博维克 著

张 勇 马东玺 译



国防工业出版社

National Defense Industry Press

现代图像质量评价

MODERN IMAGE QUALITY ASSESSMENT

[加]王周 [美]阿兰·博维克 著
张勇 马东玺 译

国防工业出版社

·北京·

著作权合同登记 图字:军-2014-63号

图书在版编目(CIP)数据

现代图像质量评价 / (加)王周, (美)博维克
(Bovik, A. C.) 著; 张勇, 马东玺译. —北京: 国防工
业出版社, 2015. 1



原名原文: Modern image quality assessment

ISBN 978-7-118-09749-8

I. ①王... II. ①博... ②张... ③马... III.

①数字图像处理 IV. ①TN911.73

中国版本图书馆 CIP 数据核字(2014)第 244652 号

Original English language edition published by Morgan and Claypool publishers

Copyright © 2006 Morgan and claypool Publishers

All Rights Reserved Morgan and Claypool Publishers

※

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

北京京华虎彩印刷有限公司印刷

新华书店经售

*

开本 880 × 1230 1/32 印张 4 1/8 字数 103 千字

2015 年 1 月第 1 版第 1 次印刷 印数 1—2000 册 定价 49.00 元

(本书如有印装错误, 我社负责调换)

国防书店: (010) 88540777

发行邮购: (010) 88540776

发行传真: (010) 88540755

发行业务: (010) 88540717

前 言

尽管围绕图像质量评价这个课题的研究已经长达 40 多年,但是直到近年也鲜有相关著作问世。出现这种情况的原因并不是人们没有这个方面的需求,或对这个课题不感兴趣。事实上,大多数图像处理算法和设备都在致力于保持或提高供人类视觉欣赏的数字图像的外观质量。

传统意义上的图像质量往往由主观评价获得。主观评价获得的实验结果虽然可信度较高,但实际应用起来耗时长、费用高。因此,作为一本图像质量客观评价的著作,本书的目的是提供能够自动预测图像感知质量的计算模型。在图像质量客观评价领域,Mannos 和 Sakrison 进行了具有开创性的工作,他们提出的图像保真度判据,将人类视觉灵敏度作为一个空间频率函数加以考虑^[1]。其他早期的重要工作被收录于 Watson 编著的一本书中^[2]。然而,尽管人们公认图像质量评价领域的研究很重要,但是在过去的几年里,该领域还是没有得到应有的重视。

长期以来,图像质量评价一直扮演着一种矛盾的角色,因为它既像是图像处理工程师和视觉科学家的“圣杯”,又像是研究开发工作的“无人区”。实际上,在我们撰写这本书的时候,近期的互联网搜索结果显示:与“图像复原”相关的文章数量是与“图像质量评价”相关的文章数量的 100 倍,“图像增强”的文章数量则是“图像质量评价”文章数量的 400 倍。为什么会出现这种情况呢?图像质量评价似乎应该是开发图像复原和增强算法的必要组成部分,难道图像质量评价是一个如此令人畏缩的、无法解决的难题吗?以至于研究者

不愿在这里浪费他们的努力。

人类观察者有一点能做到,就是能够持续准确地判断图像质量(至少在一定范围内)。但是,我们借助当今的软件和计算工具,能够效仿或者至少能够接近人类的这种视觉功能吗?要达到这一点,肯定有很多困难。要解决这个问题就必须将图像质量与人类感知的图像质量匹配起来——这就要求我们要了解人类视觉在这方面的主要功能——即使达不到神经元水平的功能,至少也应该达到能对“黑盒”系统的输入与输出进行描述的水平。

令人激动的消息是这样的模型已经出现,尽管它们还很简单,但还是能够使用这些模型创建与人类视觉功能相一致的图像质量评价算法。技术进步及互联网上随处可见的数字图像和视频引发了人们对数字图像的关注,虽然在过去的5年里,对图像质量评价领域的关注度和研究呈现加速发展态势,但它在总体上并没有与数字图像关注度的快速增长保持一致。图像质量评价领域不再是“无人区”,也不是在一般情况下均能模仿人类视觉功能的“圣杯”,我们可能还没有掌握它,但已经在地平线上看到了它。

写这本书有三个目的:一是介绍图像质量评价的基础,解释相关工程问题;二是通过描述基于不同假设的解决工程问题的主要算法,在现有条件下对图像质量评价加以最宽泛的解释,将根据“真实”图像或未失真图像、失真类型和利用的人类视觉的相关知识,对现有图像质量评价算法进行分类;三是通过介绍近来一些评价性能优良、概念新颖的算法和事例,提出未来研究的新方向。这些新算法和以往的明显不同,在今后的研究中还有待进一步完善。

细心的读者可能很快就注意到本书忽略了一些相关内容,如彩色图像质量评价和视频图像质量评价这两个重要话题。以上两种图像类型要求与视觉过程紧密联系的细致建模,代表了研究的新方向。本书没有讨论彩色图像质量评价,这是因为对颜色空间、色彩传递和显示以及色彩感知等相关内容的探讨还需要另外一本书才能详尽描述。遗憾的是,尽管彩色图像质量评价很有必要,但是对这个崭新领

域的相关内容的了解还微乎其微,因此目前写一本涵盖彩色图像质量评价的书可能还为时尚早。同样,视频质量评价目前也处于发展的初期阶段,主要原因是:与静态图像相比,人们对动态图像的视频失真建模及感知建模仍显滞后。然而,视频质量评价是本领域最富有开创性的研究方向之一,其衍生的研究方向比比皆是,如处处存在的数字电视视频流、互联网、新兴的数字影院等。我们期待视频质量评价领域在不久的将来取得重大进展。

本书面向广泛的读者群,包括对图像质量评价现状感兴趣的大学生,研发图像/视频质量评价系统的专业工程师,以及对图像质量评价新算法感兴趣,或者使用现有算法设计或优化其他图像处理应用的理论研究者等。

作为作者,我们很荣幸能够参与到由 Morgan 和 Claypool 出版社创立的创新出版项目中。在当今数字网络时代,我们的研究成果和纸质书籍能代表研究领域的先进技术,我们对此倍感欣慰。面对出版业的变迁,Morgan 和 Claypoo 迈出了关键而大胆的一步,感谢他们为我们做出的努力。

还要特别感谢 Claypool 在编写本书时表现出的非凡不懈的耐心和温和的劝说。相信 Morgan 和 Claypool 出版社有美好的未来。

同时也要感谢 Eero Simoncelli 博士、Hamid Sheikh、Ligang Lu、Junqing Chen 和 Scott Acton,他们仔细研读本书,并就本书进行了多方面富有见解的评论。

Zhou Wang
Alan C. Bovik

目 录

第 1 章 绪论	1
1.1 主观和客观图像质量评价方法	1
1.2 MSE 存在的问题	3
1.3 客观图像质量评价方法分类	10
1.3.1 全参考、无参考和部分参考图像质量评价方法	10
1.3.2 面向通用和特定应用的图像质量评价方法	11
1.3.3 自底向上和自顶向下的图像质量评价方法	12
1.4 本书的组织	13
第 2 章 自底向上全参考图像质量评价方法	14
2.1 基本思想	14
2.2 人类视觉系统	15
2.2.1 早期 HVS 原理	15
2.2.2 心理学 HVS 特征	19
2.3 误差视觉方法架构	23
2.3.1 预处理	23
2.3.2 通道分解	25
2.3.3 误差标准化	26
2.3.4 误差汇聚	28
2.4 图像质量评价算法	28
2.4.1 Daly 模型	28

2.4.2	Lubin 模型	29
2.4.3	Safranek – Johnson 模型	29
2.4.4	Teo – Heeger 模型	30
2.4.5	Watson 的 DCT 模型	30
2.4.6	Watson 小波模型	31
2.5	讨论	32
2.5.1	质量定义问题	32
2.5.2	阈值问题	32
2.5.3	自然图像复杂性问题	33
2.5.4	相关性去耦合问题	33
2.5.5	感知交互问题	34
第 3 章	自顶向下的全参考图像质量评价方法	35
3.1	基本思想	35
3.2	结构相似性方法	37
3.2.1	结构相似性和图像质量	37
3.2.2	空域结构相似性指标	38
3.2.3	复小波域结构相似性指标	48
3.2.4	结构相似性指标的说明	55
3.3	信息理论方法	55
3.3.1	信息保真度和图像质量	55
3.3.2	视觉信息保真度方法	56
3.3.3	信息理论指标的说明	63
3.4	讨论	63
第 4 章	无参考图像质量评价	66
4.1	基本思想	66
4.2	块压缩图像所用的无参考图像质量评价	68
4.2.1	空域方法	69

4.2.2 频域方法	73
4.3 小波图像压缩的无参考图像质量测量	78
4.4 讨论	85
第5章 部分参考图像质量评价	87
5.1 基本思想	87
5.2 基于自然图像统计的小波域部分参考评价方法	91
5.3 讨论	99
第6章 展望.....	102
6.1 总结	102
6.2 未来发展方向	105
参考文献.....	111

第 1 章 绪 论

21 世纪初期,作为信息表达与交流的工具,数字图像获得了广泛应用。一部介绍数字图像处理与传输理论、算法和应用主要进展的重要书籍^[3]出版了。本书以较大篇幅探讨了如何改善和保持所处理图像的外观问题。然而,不管是经过处理的数字图像,还是未经处理的数字图像,其质量仍难称完美。在图像采集、分析、传输、处理和重建过程中,都很容易发生失真现象。为了保持、控制和增强图像质量,对于图像的采集、管理、传输和处理系统来说,能够识别和量化图像质量等级就显得非常重要。为达到这一目的,有必要开发高效的图像质量自动评价系统。目前,图像质量评价领域还处于新兴阶段,要能够取得实质进展,还有待新的人类视觉模型和自然图像结构数据模型的出现。

1.1 主观和客观图像质量评价方法

在大多数图像处理应用程序中,人类是最终的接收者,所以最可靠的图像质量评价方法应该是主观评价方法。实际上,采用大量观察者评分的“平均意见分数”(Mean Opinion Score, MOS)方法很久以来都被认为是对图像质量进行评价的最佳方法。然而,MOS 方法费用昂贵且速度太慢,难以实际应用。

图像质量客观评价研究的目的是设计计算模型,该模型能够自动精确地预测和感知图像质量。用“预测”这个词,是因为除非和人类主观评价相吻合,否则某种算法提供的与质量相关的数值评价是

没有用的。换言之,算法应该能够对达到人类观察者报告的平均水平的图像质量进行预测。

客观图像质量评价方法显然在很多应用中都有巨大潜力。

第一,在图像质量控制系统中,客观图像质量评价方法可用于对图像质量的监控。例如,通过质量评价标准,图像采集系统能够监控图像质量并自动调整而获得最好的图像质量数据。网络视频服务器能够检查传送到网络上的数字图像的质量,来控制 and 配置数字流资源。随着最近网络视频资源的海量增长,这一应用有着重要意义。

第二,客观图像质量评价方法可对图像处理算法和系统进行评估。例如,在使用数码相机拍摄时,如果可选用几种降噪和恢复算法提高图像质量,则需要应用质量评价标准来判断哪一种算法能达到最佳效果。

第三,客观图像质量评价方法能够嵌入到图像处理和传输系统中,实现整个系统和参数配置的优化。例如,在视觉传输系统中,图像质量方法有助于编码器中前置滤波和位匹配算法的设计优化,以及实现解码器中重建、纠错和后置滤波的设计优化。

在设计和选择图像质量评价方法时,根据应用场景的不同,常需要在精度和复杂度之间进行权衡。例如,如果一个客观系统能够完全仿真人类视觉系统(Human Visual System, HVS)的所有相关方面,包括其内建的环境知识,那么它就能够精确地预测图像质量。不过,由于其复杂程度,对 HVS 以及环境模型的了解还很有限。随着人们在这些方面知识的增加,有望开发出非常接近于人类特性的图像质量评价系统。

然而,这种图像评价系统具有知识复杂性的特点,使得在将来实施该系统的过程可能会很复杂,使其难于应用在图像处理的算法和系统中。另外,还是有可能用简单易行的方法实现优异性能,在后面的几章中将介绍一些与先前技术相比,表现出色的此类系统。

历史上,图像质量评估方法大多数都基于像均方差(Mean Squared Error, MSE)这样的较为简单的数学方法。这在很大程度上

是因为 HVS 和自然图像结构数据方面知识的缺乏,以及对这些方法进行分析和计算比较简单,在进行设计优化时能够很方便地使用。

然而,相对于人的主观质量评价来说,这类系统的预测性能还比较差。作为理论算法设计的分析标准,这些质量评价方法具有重要的作用,但是如果用于评价真实图像的质量,无论对于处理过的图像还是没有处理过的图像,其作用甚微。实际上,直到最近 10 年,图像质量评价领域还一直处于停滞状态。由于没有新的人类视觉感知模型(对于图像或者图像构成、自然图像结构和自然场景统计等的感知),图像质量评价的研究乏善可陈,还只是图像工程和视觉领域中的一个“蛮荒地帯”。

1.2 MSE 存在的问题

MSE 是既简单又古老(不幸的是 MSE 还在广泛应用)的客观图像质量评价方法。此处,从 MSE 的定义和基于 MSE 的图像质量评价方法谈起,同时也解释为什么这个工具用于图像质量评价系统很差劲,虽然它用得很普遍。

令 $X = \{x_i | i = 1, 2, \dots, N\}$ 和 $Y = \{y_i | i = 1, 2, \dots, N\}$ 表示两幅待比较的图像,其中 N 是图像采样数(像素), x_i, y_i 分别是图像 X 和 Y 的第 i 次采样。注意:这里的排列顺序并不表示像素之间的空间位置关系,而是将它们按照一维矢量排序。由于 MSE 是用一维表达式严格定义,显然 MSE 没有包含图像中的任何空间信息。而这些空间信息对测量图像质量很有价值。MSE 不适用于图像质量评价还有更重要的原因,分析如下:

假设 X 是一幅质量完美的“原始图像”, Y 是一幅准备进行质量评估的“失真图像”。那么,MSE 和另一个相关的常用的质量评价峰值信噪比(Peak Signal - to - Noise Ratio, PSNR),分别定义为

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (1.1)$$

和

$$\text{PSNR} = 10 \lg \frac{L^2}{\text{MSE}} \quad (1.2)$$

式中: L 为图像像素亮度允许的动态范围, 对于 8 位/像素灰度图像, $L = 2^8 - 1 = 255$ 。

通过采用 l_p 范数或者闵可夫斯基判据, 可将 MSE 概括为一个为图像质量量测指标, 即

$$E_p = \left(\sum_{i=1}^N |x_i - y_i|^p \right)^{1/p}, p \in [1, \infty) \quad (1.3)$$

$p = 1$ 时, 可得出平均绝对误差; $p = 2$ 时, 得到 MSE 的平方根; $p = \infty$ 时, 得到最大绝对差分, 即

$$E_\infty = \max_i |x_i - y_i| \quad (1.4)$$

以上这些指标都比较容易计算。其中, MSE 常作为算法优化的最便利的工具, 因为它是可微的, 而且当它与线性代数的工具结合时, 对于实际问题一般可得到封闭解。另外, MSE 具有清晰的物理意义, 即误差信号 (定义为两幅待比较图像的差分信号) 的能量。能量可通过如傅里叶变换等正交变换 (或一元变换) 得到。这些就是 MSE (或 PSNR) 在图像处理、传输和许多其他信号处理领域的文献中广泛探讨的原因。

然而, 长期以来 MSE 方法一直因为对感知图像质量较差的相关性受到非议。如图 1.1 所示, 其中“原始”的爱因斯坦图像 (图 1.1(a)) 经过不同的失真处理: 平均亮度改变 (图 1.1(b))、对比度拉伸 (图 1.1(c))、脉冲噪声 (图 1.1(d))、高斯白噪声 (图 1.1(e))、模糊 (图 1.1(f))、JPEG 压缩 (图 1.1(g))、空间平移 (图 1.1(h))、空间缩放 (图 1.1(i)) 和旋转 (图 1.1(j))。我们注意到, 其中图 1.1(b) ~ (g) 所示图像与原始图像的 MSE 几乎相等, 但是这些图像的视觉质量大大不同。另外可以发现, 图像经过很小的几何修正 (图 1.1(h) ~ (j)) 后, 其 MSE 值就可能非常大, 而主观评价图像质量却有所降低。

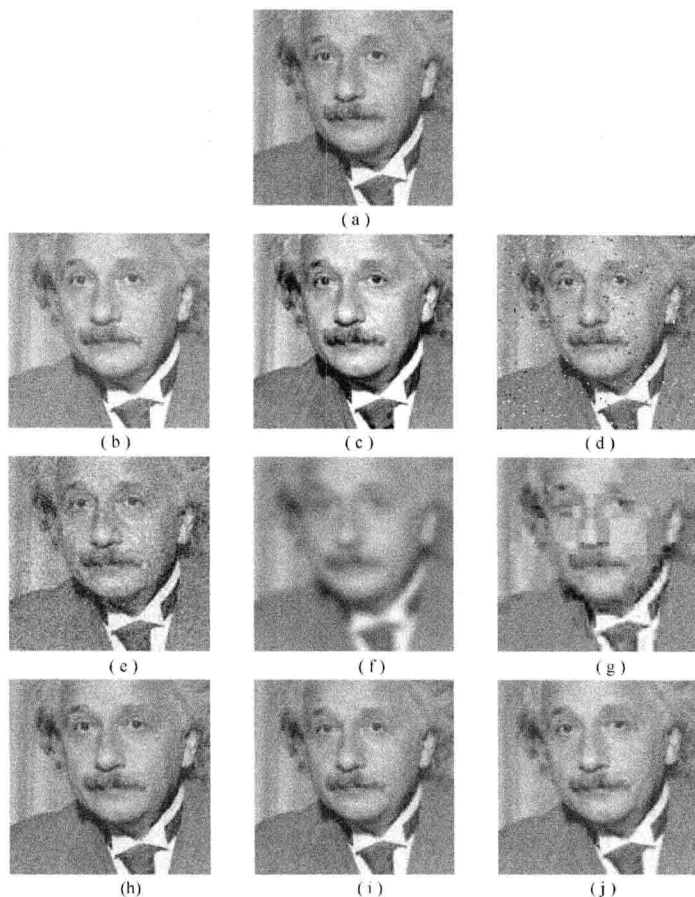


图 1.1 施加不同失真类型的爱因斯坦图像^①

(a) 原始图像; (b) 平均亮度平移, $MSE = 309$; (c) 对比度增强, $MSE = 306$;
 (d) 脉冲噪声污染, $MSE = 313$; (e) 高斯白噪声污染, $MSE = 309$; (f) 模糊, $MSE = 308$;
 (g) JPEG 压缩, $MSE = 309$; (h) 向左空间平移, $MSE = 871$;
 (i) 尺度缩放(放大), $MSE = 694$; (j) 旋转(逆时针), $MSE = 590$ 。

^① 图 1.1(b) ~ (g) 的 MSE 值几乎相同, 但视觉质量差异明显, 同时, 图 1.1(h) ~ (j) 的 MSE 表现出对空间平移、尺度和旋转的高度敏感性。

通过上述例子,人们很自然会提出疑问,“MSE 有什么问题?”最普遍的观点是 MSE 难以反映人类视觉系统感知图像的方式。目前已经发现在哺乳动物的主要视觉皮层中,输入图像表示方式与诸如 MSE 的像素域方式具有很大不同。而且,HVS 的许多重要的生理和心理特征不能用 MSE 解释。第 2 章将详细描述 HVS 特征及如何利用这些特征设计图像质量评价算法。本节从不同的角度来研究这个问题,这也许能够得到基于 l_p 范数的数学特性更为直接的答案。

即使大多数学者并未提及,实际上选择用 l_p 范数预测感知图像质量的前提,是已经做出了以下令人质疑的假设:

(1) 感知图像质量与图像信号采样的空间关系无关。因此,改变图像信号采样的空间排列不影响失真的测量。

(2) 感知图像质量与图像信号和误差信号间的关系无关。因此,对于同样的误差信号,不管隐含的图像信号是什么,失真测量保持一致。

(3) 感知图像质量仅取决于误差信号的大小。因此,改变误差信号采样的符号不影响失真测量。

(4) 所有信号采样在感知图像质量中同等重要。

遗憾的是,这些假设没有一条能满足(甚至粗略地满足)感知图像质量评价。图 1.2 ~ 图 1.5 可证明这一点。

在图 1.2 中,通过向“原始图像”(a)中添加独立的高斯白噪声得到图像(b)。图像(c)中,信号采样值与图像(a)保持不变,但是采样的空间顺序改变了(通过分类过程)。图像(d)通过产生图像(c)同样的重排序过程,由图像(b)得出;换言之,将用于图(a)→(c)的同样的坐标变换用于图(b)→(d)。不管选择的指数 p 是多少,图像(a)、(b)与图像(c)、(d)之间的 MSE 和范数 l_p 完全相同。然而,图像(d)的噪声明显比图像(b)大。这个例子说明了范数 l_p 未考虑信号采样之间的相关性(排序、模式等)。这与自然图像信号具有高度结构性这一事实是严重矛盾的,信号采样的顺序和模式携带了视觉场景中目标个体的绝大多数信息。

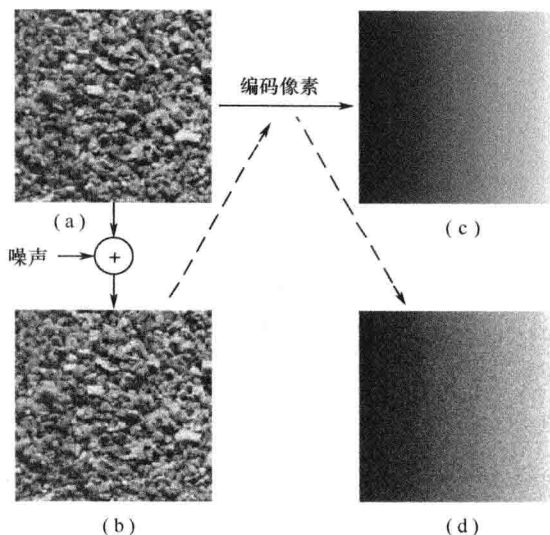


图 1.2 利用 MSE 和 l_p 范数进行图像质量预测的失败案例^①

(a) 原始图像; (b) 施加独立高斯白噪声的失真图像;

(c) 重新排序图像(a)的像素(以亮度值排序);

(d) 按照图像(c)同样的重新排序过程重新排序图像(b)的像素。

图 1.3 中,将同样的误差信号加到“原始图像”(a)、(c)中,产生的误差信号和图像(a)完全相关。失真图像(b)、(d)关于它们的原始图像都有完全相同的 MSE 和范数 l_p (不管 p 如何取值),但是图像(d)的视觉失真明显比图像(b)严重。这说明图像信号和误差信号的相关性(和依赖性)对感知图像失真有明显的影响,这一重要的性质也被范数 l_p 完全忽略了。

图 1.4 展示了两幅失真图像从原始图像产生的过程。通过对所有像素加一个常数可得到第一幅失真图像,除常数的符号随机地选择为正或负外,第二幅图像可以用同样的方法获得。但是可以发现,

① 图像(a)和(b)以及图像(c)和(d)之间的 MSE 和范数 l_p (对于任意选择的 p) 完全相同,但是图像(d)的噪声比图像(b)更大。

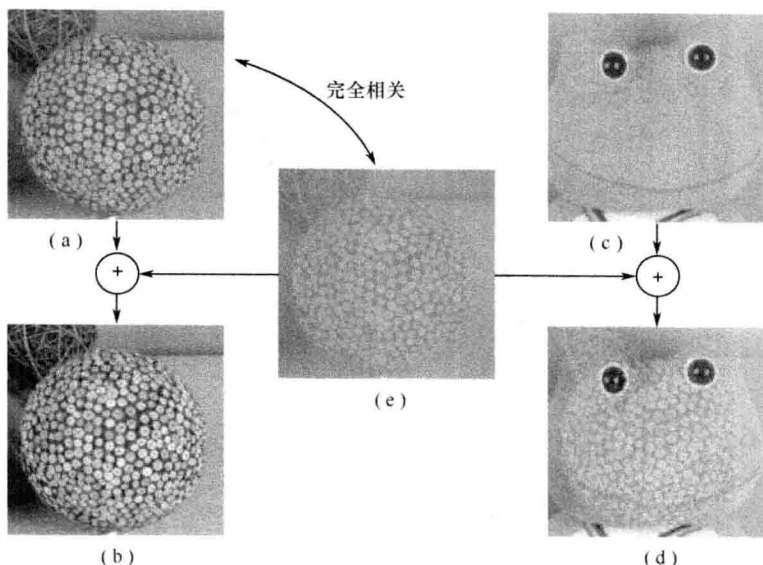


图 1.3 利用 MSE 和 l_p 范数进行图像质量预测的失败案例^①

(a)、(c) 原始图像；(b)、(d) 施加同样的误差图像(e)后的失真图像。

两幅失真图像的视觉质量区别非常大。然而,范数 l_p 完全忽略了信号的影响,不管 p 取何值,原始信号和两幅失真图像之间的范数 l_p 却完全相同。

图 1.5 中,失真图像(b)是通过向原始图像(a)叠加独立的高斯白噪声得到的。可以发现,感知的噪声水平沿着空间变化明显。脸部区域的噪声显得特别严重,但是在纹理区域则十分地微弱,可以忽略。然而,误差信号(c)沿空间具有一致的能量分布。既然所有的图像像素在所有的范数 l_p 的公式中被同样地处理,与空间相关的质量变化就不能在这些方法中得到反映。

以上讨论是 MSE 和范数 l_p 在空域中的应用,在应用于多尺度和

^① 误差图像(e)和图像(a)完全相关;图像(a)和(b)以及图像(c)和(d)之间范数 l_p (对于任意选择的 p) 相同,但图像(d)的视觉失真比图像(b)严重得多。