

全参考图像质量评价综述*

褚江, 陈强, 杨曦晨

(南京理工大学计算机科学与工程学院, 南京 210094)

摘要: 图像质量评价是图像处理领域内一项很有意义的研究课题。客观图像质量评价方法可分为全参考评价方法、半参考评价方法和无参考评价方法, 目前全参考评价方法较为成熟, 而半参考和无参考评价方法则处于初级阶段, 远远达不到参考评价方法所能达到的效果。对全参考评价方法进行综述。首先简要地介绍了各种类型的评价方法, 其次详细地介绍了 PSNR、SSIM、MSSIM、IFC、VIF、FSIM 等几种典型的全参考图像质量评价方法, 然后在 LIVE 和 TID2008 数据库上进行实验, 对这几种全参考方法进行对比、分析, 最后探讨图像质量评价研究的发展趋势。

关键词: 全参考图像质量评价; 结构相似性; 自然场景分析; 特征相似性

中图分类号: TP391

文献标志码: A

文章编号: 1001-3695(2014)01-0013-10

doi: 10.3969/j.issn.1001-3695.2014.01.003

Review on full reference image quality assessment algorithms

CHU Jiang, CHEN Qiang, YANG Xi-chen

(School of Computer Science & Engineering, Nanjing University of Science & Technology, Nanjing 210094, China)

Abstract: Image quality assessment is of fundamental importance to numerous image processing applications. Objective image quality metrics can be classified into three categories: full reference, semi-reference and no-reference. Currently, full reference metrics can achieve satisfied performance, while semi-reference and no-reference metrics are in their preliminary stage. This paper gave an overview of the full reference image quality assessment. Firstly, it introduced these methods briefly. Secondly, it described several important full reference image quality assessment algorithms in detail, such as PSNR, SSIM, MSSIM, IFC, VIF and FSIM. Then it compared the performance of these methods in the LIVE database and TID2008 database. Finally, this paper summarized the trends of future research on image quality assessment.

Key words: full reference image quality assessment; structure similarity; natural scene statistics; feature similarity

0 引言

数字图像在获取、处理、压缩、存储和传输过程中会经各种类型的退化, 这些退化会影响图像的视觉效果。一般来说, 图像质量评价方法分为主观评价方法和客观评价方法^[1]。图像质量评价的目的就是评价图像质量的好坏, 最好也最直接的方法就是直接让观察者观看, 这就是主观评价方法, 但是这种方法不能应用于实时系统中, 因此, 图像质量评价的主要目标就是设计出客观的图像质量评价方法, 使之与人的视觉感知保持一致。

根据能够得到的参考图像(原始图像)信息的多少, 客观图像质量评价方法通常可以分为三类, 分别为全参考图像质量评价、半参考图像质量评价和无参考图像质量评价^[2]。全参考图像质量评价方法可以得到参考图像的所有信息, 通过把失真图像和参考图像进行比较来评价图像质量, 全参考评价方法是目前为止客观图像评价方法中最可靠的方法。无参考图像质量评价方法不能得到参考图像的任何信息, 只对失真图像进行评价。半参考图像质量评价方法介于两者之间, 只能获得参考图像的部分信息用于评价图像质量。本文主要讨论全参考

评价方法。

最简单的全参考评价方法为均方误差(MSE)和峰值信噪比(PSNR)^[3], 它们直接对参考图像和失真图像的像素灰度值进行计算, 计算简单、意义明确, 但是不能很好地与人的主观感受保持一致^[4-5]。随着对人类视觉系统(HVS)认识的深入, 人们开始利用人类视觉系统的某些特性进行评价^[6-7], 出现了基于HVS的模型, 其中有代表性的两个模型是 Daly visible differences predictor 和 Sarnoff visual discrimination model^[8], 它们主要是通过生理和心理学实验获得人类视觉系统的前端特征构建视觉模型, 从而模拟人的视觉来对图像质量进行客观评价, 典型的方法有 NQM^[9]和 VSNR^[10], 但这些方法效果并没有得到满意的结果。Wang 等人^[2, 11]提出了结构相似性理论(SSIM), 对参考图像的亮度、对比度和结构信息进行比较, 得到了较好的结果。他假设 HVS 擅长提取场景中的结构信息, 通过评测失真图像的结构信息的退化程度对图像进行评价, 得到了广泛使用。许多学者在 SSIM 的基础上进行了改进, 如 Wang 等人^[12]提出了多尺度结构相似性(MSSIM), 得到了比单一尺度更好的结果; Li 等人^[13]提出了 3-SSIM, 把 SSIM 中对整幅图像的亮度、对比度和结构的比较, 改为在边缘、纹理和平滑

收稿日期: 2013-05-27; 修回日期: 2013-07-08 基金项目: 国家自然科学基金青年学者基金资助项目(60805003); 南京理工大学自主科研专项基金资助项目(2011ZDJH26)

作者简介: 褚江(1988-), 男, 江苏连云港人, 硕士研究生, 主要研究方向为图像质量评价(chujiang131@163.com); 陈强(1979-), 男, 副教授, 博士, 主要研究方向为图像处理、分析与识别。

区域分别计算并赋予不同的权重,得到最终的评测结果;Chen 等人^[14]提出了基于梯度的结构相似性(GSSIM),把 SSIM 中对对比度和结构的比较改为对参考图像和失真图像的梯度图像进行计算。其他的对 SSIM 的改进有:朱里等人^[15]提出的基于结构相似性的图像质量评价方法;王涛等人^[16]提出的基于内容的图像质量评价测度;黄小乔等人^[17]提出的基于色差的均方误差与峰值信噪比评价彩色图像质量评价方法;任雪等人^[18]提出的基于局部特征的图像质量评价方法;Wan 等人^[19]提出的基于感兴趣区域的方法;Aja-Fernandez 等人^[20]提出的基于局部对比度的质量评价方法;孔繁镔^[21]提出的一种把 HVS 和相似性度量结合起来的图像质量评价测度;叶盛楠等人^[22]提出的基于结构信息提取的图像评价方法。

Sheikh 等人提出了自然场景分析(NSS)的方法 IFC^[23]和 VIP^[24],它们通过评测参考图像和失真图像的共同信息的多少来评价失真图像的质量,得到了较好的结果。Liu 等人^[25]把相位一致性(PC)用于图像质量评价,相位一致性的定义可以参考文献[26]。相位一致性信息的最大特点就是它不依赖于图像的亮度 and 对比度,能够较好地表征图像的结构信息,但实验结果并不太好,因为亮度和对比度对图像的质量的影响有时并不能忽略。Zhang 等人^[27]利用相位一致性信息的特性,提出了特征相似性方法(FSIM),选取了相位一致性信息和梯度信息作为它的两个特征,得到了较好的结果。

在实际应用中,往往不能获得完美的原始图像,如在多媒体通信中,接收端很难获得原始图像的全部信息,也就大大限制了全参考图像质量评价方法的应用。因此学者们提出了半参考图像质量评价模型,即利用原始图像的部分特征信息来评价失真图像质量的一种方法。典型的半参考图像质量评价方法有基于多尺度的方法,如 Gao 等人^[28]提出的基于多尺度几何分析的方法, Lu、孙玉宝等人^[29,30]提出的基于 Contourlet 的方法, Li 等人^[31]提出的基于小波结构的方法;基于降质统计模型的方法^[32,33];基于谐波强度的方法^[34];基于小波域自然图像统计模型的方法^[35,36];基于特征嵌入的方法^[37,38]。

在很多应用中,无法得到参考图像的任何信息,这就需要无参考图像质量评价方法。无参考评价方法已成为近几年的研究热点,因为它们更加实用。大多数的无参考评价方法主要针对某一特定类型的失真来评价,如针对图像模糊的方法^[39,40]、针对 JPEG 和 JPEG2000 的方法^[41~45]。不依赖失真类型的方法大多是对 DCT 系数的统计特征进行建模^[46~48]或者是先对参考图像的统计特征进行训练,然后再进行评价^[49~53]。

1 方法介绍

1.1 传统的方法

对于 $M \times N$ 的图像,均方误差(MSE)、峰值信噪比(PSNR)和归一化最小平方误差(NLSE)的定义如下:

$$MSE = \frac{\sum_{m=1}^M \sum_{n=1}^N [R(m, n) - I(m, n)]^2}{M \times N} \quad (1)$$

$$NLSE = \sqrt{\frac{\sum_{m=1}^M \sum_{n=1}^N [R(m, n) - I(m, n)]^2}{\sum_{m=1}^M \sum_{n=1}^N [R(m, n)]^2}} \quad (2)$$

$$PSNR = 10 \log_{10} \left(\frac{L^2}{MSE} \right) \quad (3)$$

其中: $R(m, n)$ 为参考图像在坐标为 (m, n) 处的灰度值; $I(m, n)$ 为失真图像在 (m, n) 处的灰度值; L 为峰值信号,对于 8 位的灰度图像来说 $L = 2^8 - 1 = 255$ 。

$n)$ 为失真图像在 (m, n) 处的灰度值; L 为峰值信号,对于 8 位的灰度图像来说 $L = 2^8 - 1 = 255$ 。

MSE 和 PSNR 的优点是计算简单,缺点是它们单纯地计算图像之间灰度值的差异,并没有考虑像素之间的结构关系。它们将图像的所有像素点对人眼提供的信息都等同对待,而事实上人眼在观察图像的时候是有感兴趣区域的,它们的计算结果不能与图像的视觉效果保持一致。

1.2 基于结构相似度的方法

1.2.1 SSIM

Wang 等人认为结构信息反映了场景中的结构特征,同时亮度和对比度对图像的质量以及人类观察图像时提取感兴趣的区域起到很大的作用。那么它们的亮度、对比度和结构相似度分别定义如下:

假设 x, y 分别为参考图像和失真图像,则

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (4)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (5)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (6)$$

其中: μ_x, μ_y 分别为图像 x, y 的均值,反映了其亮度信息; σ_x 和 σ_y 分别为图像 x, y 的方差,反映了其对比度信息; σ_{xy} 为 x, y 的相关系数,反映了结构信息的相似度; C_1, C_2, C_3 为接近于零的正常数,防止分母为零时导致结果异常。综合这三种信息得到 x 和 y 的结构相似性为

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (7)$$

其中: α, β, γ 均大于 0,用于调整这三种信息的相对重要性。若取 $\alpha = \beta = \gamma = 1$, $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$, $C_3 = C_2/2$, $K_1 < 1$, $K_2 < 1$, L 为图像的动态范围,则式(7)可以简化为

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (8)$$

SSIM 满足以下特性:

a) 对称性, $SSIM(x, y) = SSIM(y, x)$;

b) 有界性, $SSIM \leq 1$;

c) 最大值唯一性,只有当两幅图像完全相同时, $SSIM = 1$, 否则 SSIM 的值在 0 和 1 之间。

文献[2]使用 11×11 的高斯加权窗口在参考图像和失真图像中逐像素移动,每次都在 11×11 的局部窗口中计算两幅图像的 SSIM,得到 SSIM map,最后综合所有图像块对图像进行总体评价。采用高斯加权窗是为了避免出现分块效应。由于考虑了人类视觉效果,相比于 MSE 和 PSNR 来说,SSIM 方法具有较好的相关性,计算也比较简洁,得到了广泛应用。但是 SSIM 方法是基于一个假设而建立起来的方法,即 HVS 的主要功能是提取场景中的结构信息。实际情况中,人眼的视觉信息是设计到方方面面的,包括生理和心理,所以该方法离理想的结果还有一定距离。

结构相似性给出了图像质量评价研究一个很好的研究方向,人们对结构相似性方法做出了许多改进,如 Wang 等人^[12]提出的多尺度结构相似性(MSSIM),Chen 等人^[14]提出的基于梯度的结构相似性(GSSIM),Li 等人^[13]提出的 3-SSIM。

1.2.2 MSSIM

MSSIM 把参考图像看成尺度 1,最高尺度为 M ,它是通过 $M-1$ 次迭代,每次通过对上一次迭代的结果进行低通滤波和

下采样得到的。对于第 j 个尺度,式(2)(3)中用于对比度和结构的比较记为 $C_j(x, y)$ 和 $S_j(x, y)$,亮度信息只对尺度 M 进行计算,记为 $L_M(x, y)$ 。综合多个尺度得到 MS-SSIM 为

$$\text{MS-SSIM}(x, y) = [L_M(x, y)] \cdot \prod_{j=1}^M [C_j(x, y)] \cdot [S_j(x, y)] \quad (9)$$

MS-SSIM 能够捕获跨越多个尺度的模糊,能够更好地与人类的感知保持一致,评价效果要好于单尺度的 SSIM。

1.2.3 GSSIM

Chen 等人对 SSIM 进行改进,提出了基于梯度的结构相似度(GSSIM),把 SSIM 中对对比度和结构的度量改为对参考图像和失真图像的梯度图像进行计算,分别记为 $C_g(x, y)$ 和 $S_g(x, y)$ 。即

$$C_g(x, y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (10)$$

$$S_g(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (11)$$

其中: x 和 y 为参考图像和失真图像的梯度图像

$$\text{GSSIM}(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_x \sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (12)$$

改善的效果也要好于 SSIM。

1.2.4 3-SSIM

Li 等人提出了 3-SSIM,把 SSIM 中对整幅图像的亮度、对比度和结构的比较,改为在边缘、纹理和平滑区域分别计算并赋予不同的权重,得到最终的评测结果。

该方法的计算步骤如下:

- 计算 SSIM map。
- 把参考图像和失真图像分为三个区域。
- 对不同的区域赋予不同的权重(边缘权重赋为 0.5,纹理和平滑区域为 0.25),加权计算得到 3-SSIM 的值。

把图像划分为三个区域的过程如下:

a) 使用 Sobel 算子计算参考图像和失真图像的梯度图像,确定两个阈值 $TH_1 = 0.12 \times g_{\max}$, $TH_2 = 0.06 \times g_{\max}$,其中 g_{\max} 为参考图像的最大梯度值。

b) 根据以下条件把像素点分为边缘、纹理和平滑区域。把坐标为 (i, j) 处的参考图像的梯度记为 p_1 ,失真图像在相应点处的梯度记为 p_2 。

如果 $p_1 > TH_1$ 或者 $p_2 > TH_1$,该点就被划分为边缘点;

如果 $p_1 < TH_2$ 并且 $p_2 \leq TH_1$,该点被划分为平滑区域的点;

否则,把该点分为纹理区域的点。

3-SSIM 也得到了较为满意的结果,效果要好于 SSIM 和 MSSIM。

1.3 基于自然场景分析(NSS)的方法

通常把使用高质量的捕获设备获得的视频或图像称为自然场景(natural scene),自然场景只是所有可能图像的一小部分。现实世界中大多数的退化过程都会扰乱这些统计特征,使得图像或视频信号变得“不自然”。许多学者已经研究出了一些对自然场景统计特征进行建模的方法,典型的基于 NSS 的方法有 VIF^[24] 和 IFC^[23]。

1.3.1 IFC

如图 1 所示,源图像通过退化通道后图像变得失真,可以通过衡量失真图像(输出)和源图像(输入)的共同信息的多少来评价图像质量,即为 IFC。

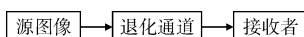


图 1 IFC 示意图

由于小波分解的子带系数既不是独立的,也不是同分布的,一个系数的邻域如果有较大的方差,它也可能有较大的方差,它们的边缘密度分布在零点处呈现类似拉普拉斯式的分布(在零点处有很强的峰值)。文中使用小波域的 GSM(Gaussian scale mixture)模型^[54]对 NSS 建模,首先对一个子带进行处理,然后再扩展到多个子带。

GSM 是一个随机场(RF),可以表达成两个独立随机场的乘积,即对于 GSM $C = \{C_i; i \in I\}$,可以表达为

$$C = S \cdot \mu = \{S_i \cdot U_i; i \in I\} \quad (13)$$

其中: i 为随机场的空间下标; $S = \{S_i; i \in I\}$ 是一个正标量的随机场; $\mu = \{U_i; i \in I\}$ 是一个高斯随机场,均值为 0,方差为 C_u ; C_i 和 U_i 是 M 维的向量。对于随机场 μ ,假设对任意 i 和 $j(i \neq j)$, U_i 和 U_j 相互独立。给定 S, C 服从正态分布,给定 S_i ,对任意 j , C_i 独立于 S_j ,给定 S, C_i 条件独立于 C_j 。

退化模型描述的是图像是如何退化的,一个易于计算的模型为

$$D = gC + V = \{g_i C_i + V_i; i \in I\} \quad (14)$$

其中: C 是参考图像一个子带的 RF, $D = \{D_i; i \in I\}$ 是退化图像相应子带的 RF, $g = \{g_i; i \in I\}$ 是衰减因子,能够捕获由于模糊造成子带信号能量的丢失, $V = \{V_i; i \in I\}$ 是加性高斯噪声,均值为 0,方差为 $C_v = \sigma_v^2 I$,能够捕获噪声对图像的影响。

假设 $C^N = (C_1, C_2, \dots, C_N)$ 表示 C 的 N 个元素, $D^N = (D_1, D_2, \dots, D_N)$ 表示 D 的 N 个元素, C 和 D 的共同信息表示为 $I(C_N; D_N)$,由于 C 和 S 的非线性依赖关系,为了便于分析,假设 S 是已知的。 C 和 D 的共同信息可以表示为

$$I(C^N; D^N | S^N) = \sum_{i=1}^N \sum_{j=1}^N I(C_i; D_j | C^{i-1}, D^{j-1}, S^N) \quad (15)$$

$$\sum_{i=1}^N I(C_i; D_i | C^{i-1}, D^{i-1}, S^N) \quad (16)$$

$$\sum_{i=1}^N I(C_i; D_i | S_i) \quad (17)$$

由于对任意的 i 和 $j(i \neq j)$, D_i 独立于 C_j ,对于给定的 S_i , C_i 和 D_i 独立于 S_j ,由于 C_i 和 V_i 都服从高斯分布,可以得到

$$I(C^N; D^N | S^N) = \sum_{i=1}^N I(C_i; D_i | S_i) = \quad (18)$$

$$\sum_{i=1}^N (h(D_i | S_i) - h(D_i | C_i, S_i)) = \quad (19)$$

$$\sum_{i=1}^N (h(g_i C_i + V_i | S_i) - h(V_i)) = \quad (20)$$

$$\frac{1}{2} \sum_{i=1}^N \log_2 \left(1 + \frac{g_i^2 S_i^2 \sigma_u^2}{\sigma_v^2} \right) \quad (21)$$

其中: $h(X)$ 表示连续随即变量 X 的微分熵,对于服从正态分布 $N(\mu, \sigma^2)$ 的 X , $h(X) = 1/2 \log_2(2\pi e \sigma^2)$ 。

把所有子带的共同信息加起来就得到了 IFC。

$$\text{IFC} = \sum_{k \in \text{subbands}} I(C^{N_k k}; D^{N_k k} | S^{N_k k}) \quad (22)$$

IFC 和其他基于 HVS 的方法相比,最大的特点是它不需要配置 HVS 的任何参数,计算上也比较简洁,得到了较满意的结果。

1.3.2 VIF

如图 2 所示, VIF 方法通过衡量两种共同信息来评价图像质量,一种信息是经由 HVS 通道(没有失真存在时)得到的输出图像和输入图像的共同信息,称为参考图像信息,另一种信息是失真通道的输入和输出的共同信息,称为失真图像信息, VIF 即为这两种信息的比值。



图 2 C 和 E 的共同信息可以衡量大脑从参考图像中
提取的信息的多少 相应地 C 和 F 的共同信息
可以衡量大脑从失真图像中提取的信息

然后对 HVS 建模,本文把 HVS 看成一个退化通道,认为图像经过 HVS 后只有部分信息被接受。为了计算上的简洁,把 HVS 的退化看成只包含噪声项,称为视觉噪声,在小波域对 HVS 噪声建模:

$$E = C + N(\text{参考图像})$$

$$F = D + N(\text{失真图像})$$

其中: $N = \{N_i; i \in I\}$ 和 $N' = \{N'_i; i \in I\}$ 是与 C 有同样维数的零均值,多元高斯 RF; E 是参考图像一个子带经由 HVS 后大脑提取的信息, F 是退化图像的一个子带经由 HVS 后大脑提取的信息,如图 2 所示。

假设 N 和 N' 独立于 U、S 和 V 对 N 和 N' 的方差建模

$$C_N = C_{N'} = \sigma_n^2 I \quad (23)$$

其中: σ_n^2 是 HVS 模型参数。

共同信息 $I(C; E)$ 衡量的是一副图像经由 HVS 输出后大脑能够提取的信息的数量,相应地 $I(C; F)$ 为退化图像经由 HVS 后大脑能够提取的信息的数量,具体的计算过程参见文献[24]。

$$I(C_N; E_N | s_N) = \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^M \log_2 \left(1 + \frac{s_i^2 \lambda_k}{\sigma_n^2} \right) \quad (24)$$

$$I(C_N; F_N | s_N) = \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^M \log_2 \left(1 + \frac{g_i^2 s_i^2 \lambda_k}{\sigma_n^2 + \sigma_v^2} \right) \quad (25)$$

这两种信息的比值即为 VIF,如式(26)所示。

$$VIF = \frac{\sum_{j \in \text{subbands}} I(C_{N_j}; F_{N_j} | s_{N_j})}{\sum_{j \in \text{subbands}} I(C_{N_j}; E_{N_j} | s_{N_j})} \quad (26)$$

其中: C_{N_j} 表示子带 j 的 N 个系数。

VIF 的最大特点是,对于线性对比度增强的图像,计算得到的 VIF 的值大于 1,而其他情况下 VIF 的值在 0 和 1 之间。这在很大程度上符合人的视觉感受,因为对比度增强的主要目的就是增强图像的视觉效果。

1.4 特征相似性(FSIM)

人眼主要是根据图像中的低级特征(如阶跃边缘、零交叉边缘等)来理解图像。基于生理和心理学研究发现,人类视觉可察觉到的特征和不同频率下傅里叶级数的相位一致性是一致的,也就是说人类感觉到的图像特征总是位于相位一致性高的点上。相位一致性可以被看做是体现局部结构信息重要性的一种无量纲。Liu 等人^[25]把相位一致性(PC)用于图像质量评价,特征相似性(FSIM^[27])选取相位一致性信息作为第一个特征。考虑到相位一致性不受对比度的影响,而对对比度会影响图像的质量,因此FSIM引入了图像梯度作为第二个特征。相位一致性和梯度信息共同合作,有效地提取人类感兴趣的特征点,基于特征相似性的图像质量评价方法也就取得了比较好的结果。

1.4.1 相位一致性

相位一致性理论假设图像中傅里叶分量级数最高的点为特征点,而并不是简单地认定灰度变化最大的点为特征点。经过生理学和物理学的证明,相位一致性可以提供一种简单而有效的方法来模拟人类视觉系统观察和感知图像的特征^[55]。

由 M_n^e 和 M_n^o 表示的 n 级偶对称和奇对称滤波器形成了一个正交组合,在 x 点处,每一对正交信号的相位都会形成一个对应的向量: $[e_n(x) \quad o_n(x)] = [g(x) \cdot M_n^e \quad g(x) \cdot M_n^o]$ n 级振幅 $A_n(x) = \sqrt{e_n(x)^2 + o_n(x)^2}$ 。假设 $F(x) = \sum_n e_n(x)$, $H(x) = \sum_n o_n(x)$ 。那么,一维信号的相位一致性可以定义为

$$PC(x) = E(x) / (\varepsilon + \sum_n A_n(x)) \quad (27)$$

其中: ε 是一个比较小的正常数。

关于双正交滤波器,一般采用 LogGabor 滤波器。LogGabor 滤波器在频域的传递函数可以定义为 $G(w) = \exp(-\log(w/w_0)^2 / (2\sigma_r^2))$,其中 w_0 是滤波器的中心频率, σ_r 则控制滤波器的带宽。

通过在垂直方向上使用扩展函数,上述的一维 LogGabor 滤波器就可以扩展为二维滤波器。使用高斯函数作为扩展函数,二维的 LogGabor 函数的传递函数如下:

$$G_2(W, \theta_j) = \exp\left(-\frac{(\log(w/w_0))^2}{2\sigma_r^2}\right) \times \exp\left(-\frac{(\theta - \theta_j)^2}{2\sigma_\theta^2}\right) \quad (28)$$

其中: $\theta_j = j\pi/J$, $j = \{0, 1, \dots, J-1\}$ 是滤波器的方向角, J 是方向数, σ_θ 是滤波器的角带宽。

通过调节 w_0 和 θ_j 以及将二维图像的 G_2 卷积,每个像素点都可以得到一个返回值,记做 $[e_{n\theta_j}(x) \quad o_{n\theta_j}(x)]$ θ_j 方向的振幅 $A_{n\theta_j}(x) = \sqrt{e_{n\theta_j}(x)^2 + o_{n\theta_j}(x)^2}$ 。 θ_j 方向的能量 $E_{\theta_j}(x) = \sqrt{F_{\theta_j}(x)^2 + H_{\theta_j}(x)^2}$ 。其中, $F_{\theta_j}(x) = \sum_n e_{n\theta_j}(x)$, $H_{\theta_j}(x) = \sum_n o_{n\theta_j}(x)$ 。那么像素点 x 的相位一致性定义为

$$PC_{2D}(x) = \frac{\sum_j E_{\theta_j}(x)}{\varepsilon + \sum_n \sum_j A_{n\theta_j}(x)} \quad (29)$$

1.4.2 梯度

梯度可以通过卷积掩模来表示, Sobel 算子^[56]、Prewitt 算子和 Scharr 算子^[57]是最常用到的计算梯度的三种方法。图像的梯度信息幅度 GM 就可以定义为 $G = (G_x^2 + G_y^2)^{1/2}$,其中 G_x 和 G_y 分别为在水平和垂直方向上计算得到的梯度值。

1.4.3 灰度图像的 FSIM 的计算

在得到相位一致性和梯度以后, PC_1 和 PC_2 表示从两幅图像中提取得到的相位一致性的值, G_1 和 G_2 则表示梯度得到的梯度信息值。将图像的质量测度分为两部分

$$S_{PC}(x) = \frac{2PC_1(x) \cdot PC_2(x) + T_1}{PC_1^2(x) + PC_2^2(x) + T_1} \quad (30)$$

$$S_{GM}(x) = \frac{2GM_1(x) \cdot GM_2(x) + T_2}{GM_1^2(x) + GM_2^2(x) + T_2} \quad (31)$$

其中: T_1 、 T_2 是常量。

将 S_{PC} 和 S_{GM} 结合起来得到参考图像和失真图像之间的相似性 $S_L(x)$ 为

$$S_L(x) = [S_{PC}(x)]^\alpha \cdot [S_{GM}(x)]^\beta \quad (32)$$

在本方法中,把 α 和 β 都设为 1。

在得到每一点 x 的 $S_L(x)$ 值之后,总的图像之间的相似度就可以得到了。但是,人类对图像上所有点的兴趣度是不一样的,如边缘点就比平滑区域更为关键。因此,相位一致性相比较而言更为重要。相位一致性的值反映了此像素点为人类视觉感兴趣的点的可能性。直观地,对于一个给定的像素点 x,如果参考图像 f_1 和失真图像 f_2 在该点的 PC 值比较大,那么该像素点就是人类感兴趣的点,因此使用 $PC_m = \max(PC_1(x), PC_2(x))$ 来调节 $S_L(x)$ 在图像质量测度中的重要性。

综上得知,参考图像和失真图像之间的综合相似性测度为

$$FSIM = \frac{\sum_{x \in \Omega} S_L(x) \cdot PC_m(x)}{\sum_{x \in \Omega} PC_m(x)} \quad (33)$$

其中: Ω 表示图像的整个空域。

1.4.4 彩色图像的 FSIM 的计算

FSIM 相对于其他图像质量评价方法还有一个优点是它可以对彩色图像进行计算。一般的图像质量评价方法都是针对灰度图像的,因为色度同样会影响到 HVS 对图像的理解。如果在 FSIM 中考虑了色度信息,则结果将会更加可靠。通过简单地扩展针对灰度图像的 FSIM 方法就可以得到针对彩色图像的 FSIM 方法。

首先将原始的 RGB 图像转换到另一个色彩空间 YIQ^[58], 该色彩空间可以将图像的亮度和色度区分开来。其中, Y 代表图像的亮度信息, I 和 Q 代表图像的色度信息,颜色空间的转换公式如下:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (34)$$

假设原始图像和失真图像分别为 f_1 和 f_2 , I_1 (I_2) 和 Q_1 (Q_2) 分别表示原始图像和失真图像的色度分量。图像色度之间的相似度可以定义为

$$S_I(x) = \frac{2I_1(x) \cdot I_2(x) + T_3}{I_1^2(x) + I_2^2(x) + T_3} \quad (35)$$

$$S_Q(x) = \frac{2Q_1(x) \cdot Q_2(x) + T_4}{Q_1^2(x) + Q_2^2(x) + T_4} \quad (36)$$

其中: T_3 和 T_4 是调节因子,因为 I 和 Q 有相似的动态范围,为了简单起见,这里假设 $T_3 = T_4$ 。因此,色度信息 $S_c(x)$ 就可以通过 $S_I(x)$ 和 $S_Q(x)$ 结合得到,如式(37)所示。

$$S_c(x) = S_I(x) \cdot S_Q(x) \quad (37)$$

最终,彩色图像的 FSIM 可以通过灰度图像的 FSIM 简单扩展得到,如式(38)所示。

$$FSIM_c = \frac{\sum_{x \in \Omega} S_L(x) \cdot [S_c(x)]^\lambda \cdot PC_m(x)}{\sum_{x \in \Omega} PC_m(x)} \quad (38)$$

其中: λ 是调节因子,可以调节色度信息在所有彩色信息中的权重。

彩色图像的计算比针对灰度图像的计算要稍微复杂一点,不过也更加符合实际需求,因为现实世界中的场景都是缤纷多彩的,并不是灰度图像。因此,基于彩色图像的质量评价要比基于灰度图像的图像质量评价更为可靠,评价结果也要好于灰度图像。

2 实验结果与分析

为了衡量评价方法的好坏,本文在 LIVE 和 TID2008 两个数据库上对 PSNR、SSIM、MSSIM、IFC、VIF、NQM、VSNR、FSIM 和 FSIMc 九种方法进行测试,其中 FSIMc 是彩色图像的特征相似度,它是针对彩色图像的评价方法,其他的方法都只针对彩色图像的亮度分量进行处理。

2.1 图像评价数据库

有六个公开的图像评价数据库,包括 TID2008^[59]、CSIQ^[60]、LIVE^[61]、IVC^[62]、MICT^[63] 和 A57^[64],其中以 LIVE 和 TID2008 最为完善。LIVE 数据库被认为是图像质量评价的标准数据库;TID2008 图像库是专为评价全参考型图像质量评价而建立的,它的优点是图像数量多、参与人员广泛,有较高的权威性和可靠性。因此本文选用 LIVE 和 TID2008 这两个数据库

进行测试。这些数据库的源图像数量、失真图像数量、失真类型数、图像类型和建库时的观察者数量如表 1 所示。TID2008 图像库的每一种失真类型只有四个失真等级,相比之下 LIVE 要多得多。因此,在对 LIVE 库做实验时,针对某一种失真类型的图像做实验就可以验证评价方法的可靠性;而在 TID2008 做实验时,针对所有图像做实验要比针对每一种失真类型做实验更加能够验证评价方法的可靠性。

表 1 图像评价数据库

数据库	源图像数量	失真图像数量	退化类型数	图像类型	观察者
TID2008	25	1 700	17	24b/p 彩色	838
LIVE	29	779	5	24b/p 彩色	161
CSIQ	30	866	6	24b/p 彩色	35
IVC	10	185	4	24b/p 彩色	15
MICT	14	168	2	24b/p 彩色	16
A57	3	54	6	8b/p 灰度	7

2.2 评价指标

为了衡量方法测试结果与主观评价之间的一致性,视频质量专家组 VQEG^[65] 提出了四个可以验证客观评价结果和主观评价结果之间的紧密程度的四个指标,它们分别是斯皮尔曼等级次序相关系数 (ROCC)、肯德尔等级次序相关系数 (KROCC)、皮尔逊线性相关系数 (LCC) 以及均方误差 (MSE)。

斯皮尔曼等级次序相关系数和肯德尔等级次序相关系数用于衡量评价结果的单调性,这两个指标只衡量数据点的等级次序,而不考虑数据点间的相对距离。为了测量其他两个指标,需要先进行回归分析,建立客观分值与主观分值 (MOS) 的非线性映射,皮尔逊线性相关系数评价的是 MOS 值与非线性回归后的客观分值的相关性;MSE 是非线性回归后的客观分值与 MOS 值的均方误差。本文使用的非线性回归方法为 Logistic 函数^[66]。当客观评价结果与主观评价结果最好时,斯皮尔曼等级次序相关系数、肯德尔等级次序相关系数和皮尔逊线性相关系数的值会很高,而均方根误差的值会很小。

假设主观评价数组为 x ,客观评价数组为 y , n 为 x 和 y 的大小,斯皮尔曼相关系数 (SROCC) 是指对两个目标数组的秩次大小作线性相关分析,通常被认为是两个对象分别经过排列后的皮尔逊线性相关系数,适用范围比较广泛。首先将 x 、 y 按照从大到小的顺序进行排列,记 x_j 和 y_j 分别为原数组中 x_i 、 y_i 在排序之后的数组中的位置,如果 x_i 、 y_i 不同,则 SROCC 可以按如下公式计算:

$$SROCC = 1 - \frac{6 \sum (x_j - y_j)^2}{n(n^2 - 1)} \quad (39)$$

肯德尔等级次序相关系数 (KROCC) 反映了分类变量的相关性,与斯皮尔曼相关系数一样适用于两个对象数组都是有序的情况。

$$KROCC = (P - Q) \frac{2}{n(n-1)} \quad (40)$$

其中: P 为和谐对的个数, Q 为不和谐对的个数, n 为样本容量。和谐对是指变量大小顺序相同的两个样本观测值,即 x 等级高低顺序与 y 等级高低顺序相同,否则称为不和谐。

皮尔逊线性相关系数 (PLCC) 常用来度量两个线性变量之间相关性的强弱,其取值范围在 $[-1, 1]$ 间,并且具有对称性,两个对象的皮尔逊线性相关系数定义为这两个对象的协方差

与两者标准差的商。

$$PLCC = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{E(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y} \quad (41)$$

与上面三个系数不同,均方根误差表示了两个对象数组之间的离散程度,值越小说明相关性越好。均方根误差又叫做标准误差,其计算公式为

$$MSE = \sqrt{\frac{\sum (x_i - y_i)^2}{n-1}} \quad (42)$$

2.3 实验结果

图3是这九种方法在LIVE库上的散点图。表2是这几种方法在LIVE库上的SROCC的值,其中加粗部分表示的是每一种失真类型(每一列)SROCC最大的3个值;表3为这些方法的运行时间;表4为这些方法在TID2008库上的SROCC的值。图4为这九种方法在TID2008数据库的散点图。在Pentium 1.86 GHz的计算机上使用MATLAB 7.0对768×512的图像进行测试,运行时间如表3所示,单位为s。从表中可以看出PSNR的计算时间最少,因为它只针对像素点对的灰度值进行比较,SSIM也比较快,IFC、VIF和FSIM要慢一些,但从表2和

4中可以看出它们的效果要好于PSNR和SSIM。

表2 几种方法在LIVE库上SROCC的值

方法	JPG2K	JPEG	WN	GBLUR	FF
PSNR	0.889 8	0.840 9	0.985 3	0.781 6	0.890 3
NQM	0.929 7	0.896 3	0.981 9	0.842 2	0.814 4
IFC	0.889 1	0.862 2	0.937 3	0.958 2	0.942 6
VIF	0.951 8	0.954 4	0.937 0	0.937 0	0.964 9
SSIM	0.931 7	0.902 8	0.962 9	0.894 2	0.941 1
MSSIM	0.954 9	0.912 4	0.937 1	0.958 4	0.931 7
VSNR	0.939 8	0.908 4	0.981 9	0.934 2	0.903 9
FSIM	0.956 9	0.913 0	0.965 2	0.971 0	0.949 7
FSIMc	0.957 4	0.913 4	0.971 6	0.971 1	0.951 5

表3 九种方法运行时间比较

方法	PSNR	SSIM	MSSIM	IFC	VIF
运行时间	0.1400	0.8120	1.4850	7.8130	7.9840
方法	VSNR	NQM	FSIM	FSIMc	
运行时间	1.6570	4.5000	7.5480	7.9070	

表4 九种评价方法在TID2008数据库上SROCC的值

	FSIM	FSIMc	MS-SSIM	VIF	SSIM	IFC	VSNR	NQM	PSNR
awgn	0.856 6	0.875 8	0.809 7	0.88	0.797 4	0.581 7	0.772 8	0.767 9	0.907 1
wagn_color	0.852 7	0.893 1	0.805 9	0.878 5	0.811 4	0.552 8	0.779 3	0.749	0.898 8
spatial corr-noise	0.848 3	0.871 1	0.819 4	0.870 3	0.827 1	0.598 4	0.766 5	0.772	0.916 3
masked noise	0.802 1	0.826 4	0.815 1	0.869 8	0.811 6	0.732 6	0.729 5	0.706 7	0.851 3
high fre-noise	0.909 3	0.915 6	0.868 3	0.907 5	0.843 1	0.736 1	0.881 1	0.901 5	0.926 9
impulse noise	0.745 2	0.771 9	0.686 8	0.833 1	0.746 2	0.533 4	0.647 1	0.761 6	0.872 5
quantization noise	0.856 4	0.872 6	0.853 9	0.795 7	0.820 8	0.591 1	0.827	0.820 9	0.869 8
blur	0.947 2	0.947 2	0.960 5	0.954 6	0.938 8	0.876 6	0.933	0.884 6	0.869 9
denosing	0.960 3	0.961 8	0.957 1	0.918 9	0.927 5	0.800 2	0.928 6	0.945	0.941 7
jpeg comp	0.927 9	0.929 4	0.934 9	0.916 9	0.898 9	0.818 1	0.917 4	0.907 5	0.871 7
jpg2k comp	0.977 3	0.978	0.973 4	0.971 3	0.887 2	0.944 5	0.951 5	0.953 2	0.813 3
jpeg trans error	0.870 8	0.875 6	0.873 7	0.858 2	0.819	0.796 6	0.805 5	0.737 3	0.751 2
jpg2k trans error	0.854 4	0.855 5	0.852 5	0.851	0.839 7	0.730 3	0.790 9	0.726 2	0.831 1
pattern noise	0.749 1	0.751 4	0.733 5	0.760 8	0.694 8	0.841	0.571 6	0.68	0.581 5
block distortion	0.849 2	0.846 4	0.760 9	0.832	0.884 9	0.676 7	0.192 6	0.234 8	0.618 8
mean shift	0.672	0.655 4	0.736	0.513 2	0.717 6	0.437 5	0.372 7	0.524 5	0.695
contrast	0.648 1	0.651	0.639 7	0.819	0.484 2	0.274 8	0.423 9	0.619 1	0.587 2

2.4 结果分析

传统的PSNR和MSE等方法有着计算简单、物理意义明确等优点,但它们仅仅是对像素点之间误差的纯数学统计,并没有考虑到像素点之间的相关性和人眼视觉系统的感知特性,不能真实地反映图像的视觉感知质量,在很多情况下不符合人的主观感受。PSNR方法认为图像的失真由于噪声的加入导致的,所以它对白噪声的图像有较好的结果,从表2、3和图3的(a4)中可以看出,但对于其他的失真类型效果并不理想。NQM和VSNR是基于噪声的方法,它们模拟HVS,建立退化模型,对噪声图像具有较好的效果,但由于对HVS认识的不足,这些方法在其他类型的失真效果并不理想。

SSIM从高层次上模拟HVS的整体功能,认为HVS的主要功能是提取场景中的结构信息,得到了较好的结果,因此用结构信息的度量作为图像感知质量的近似。SSIM的方法简单、

准确性较好,提出后受到了国内外学者的广泛关注。SSIM的缺点在于没有区别对待图像中人眼感兴趣程度不同的区域,而是将子块的结构相似度和SSIM简单地取平均而得出整幅图像的结构相似度。把整个图像的各个重要程度看成是相同的,但有时各部分重要程度并不一样,这存在一定的局限性,SSIM在评价严重模糊的图像时会产生不符合人的主观感受的结果。

MSSIM是对SSIM的改进,能够捕获跨越多个尺度的模糊,能够更好地与人类的感知保持一致,评价效果要好于单尺度的SSIM,这从图4中可以清楚地看出,MSSIM与MOS值具有较好的一致性,但MSSIM存在着与SSIM同样的问题。3-SSIM对纹理、边缘和平滑区域赋予不同的权重,在一定程度上解决了该问题,效果也要好于SSIM。研究发现,人眼对于场景中的边缘和轮廓信息更为敏感,在此基础上得到的GSSIM也得到了好于SSIM的结果。

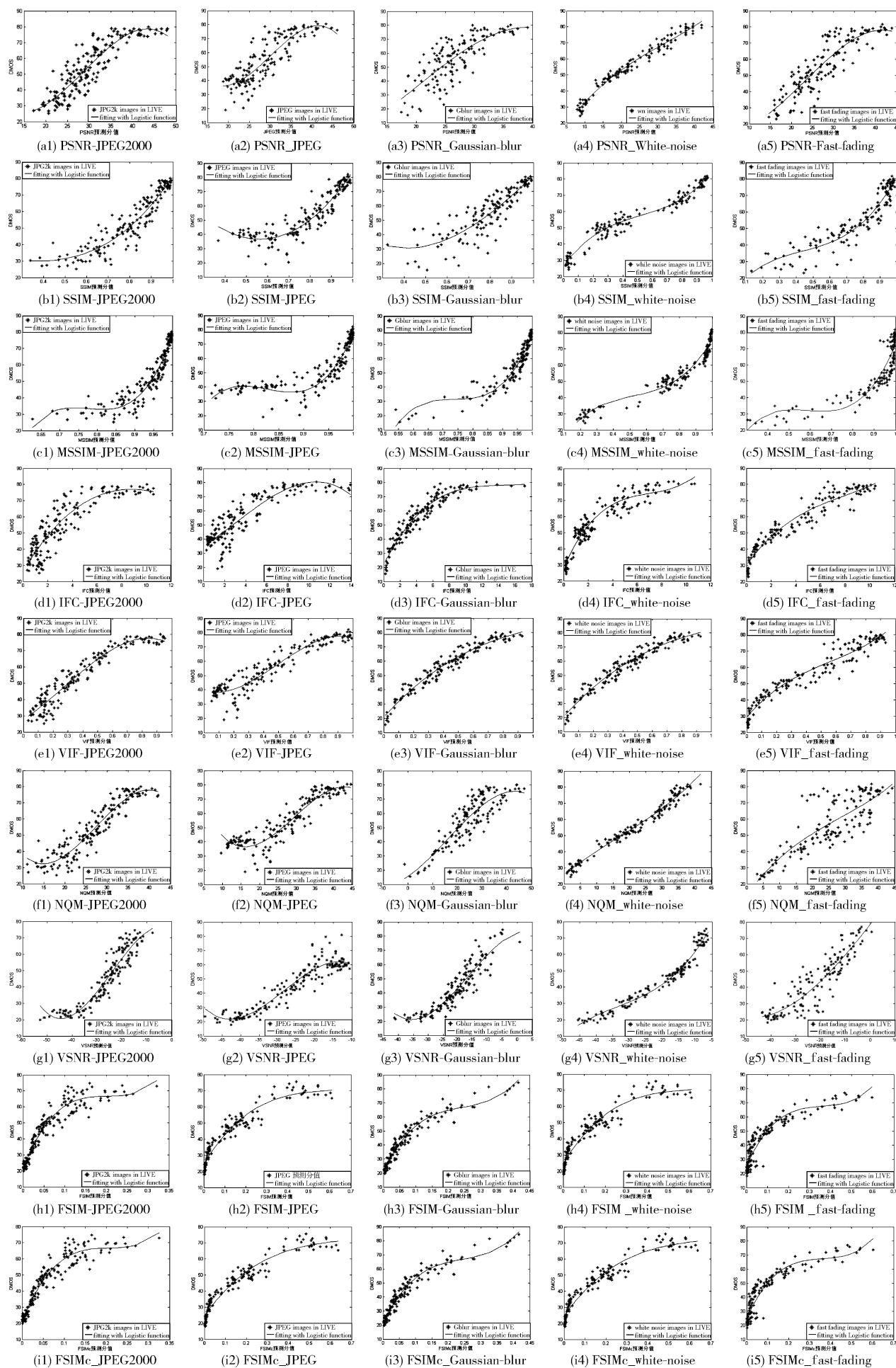


图3 九种方法在 LIVE 数据库上的评价与 MOS 值的关系曲线

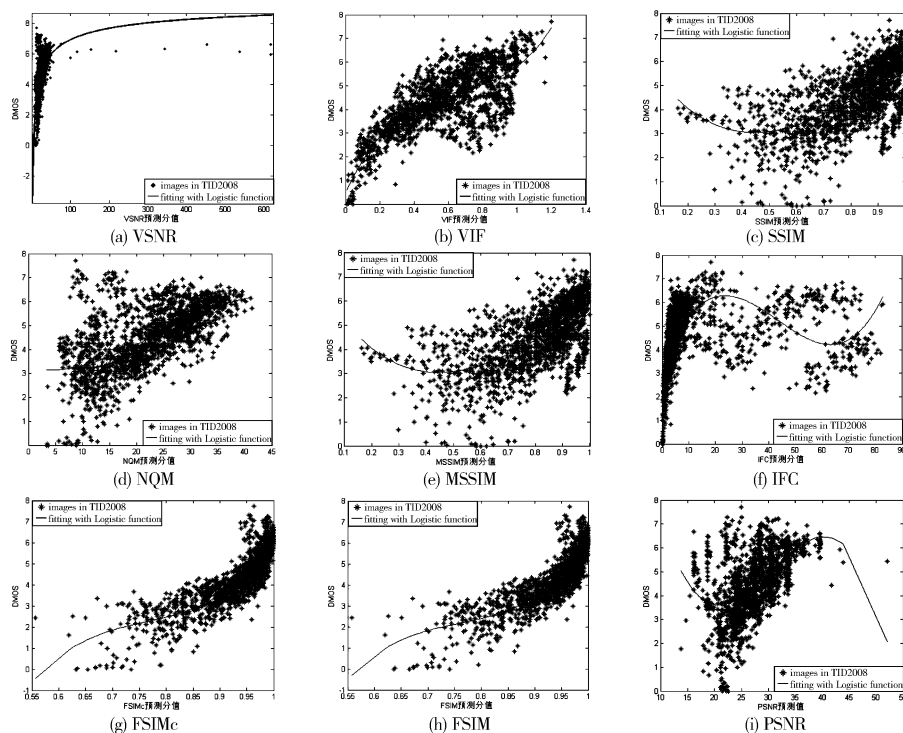


图 4 九种方法在 TID2008 数据库上的评价与 MOS 值的关系曲线

IFC 和 VIF 是基于 NSS 的方法,也需要对 HVS 建模,与其他基于 HVS 的方法相比,最大的特点是它们不需要配置 HVS 的任何参数。VIF 对噪声和模糊图像具有较好的效果。但它们的运算复杂度较高,运算时间较长,VIF 的最重要特征是对对比度拉伸的图像有很好的效果,从表 4 中可以看出,评价结果要远好于其他评价方法。

FSIM 和 FSIMc 在几乎所有失真类型上都能有较满意的结果,评价结果比较稳定,是各种方法中最好的,但是计算上要慢一些。由于现实世界是丰富多彩的,彩色图像的特征相似性 FSIMc 在各方面都要好于对灰度图像计算的 FSIM,这也符合人的视觉感受。如何减少 FSIM 的计算时间是进一步研究的主要工作。

3 结束语

本文对全参考图像质量评价方法进行综述,首先概略地介绍了图像质量评价领域的一些方法,然后详细地介绍了 PSNR、SSIM、MSSIM、IFC、VIF、FSIM 等几种典型的全参考图像质量评价方法,并使用这些方法在 LIVE 和 TID2008 数据库上进行测试,并对测试结果进行了分析和比较,讨论了这些方法的优缺点。由于全参考评价方法假设能得到参考图像的全部信息,通过把失真图像与原始图像进行比较来评价图像的质量,研究的难度要低于半参考和无参考评价方法。目前全参考评价方法比起半参考和无参考评价方法要成熟得多,也取得了较为满意的结果,对全参考图像质量评价方法的研究也有助于对半参考和无参考评价方法的设计。

笔者认为今后图像质量评价将会从以下几个方面入手进行研究:

a) 目前大多数的图像质量评价方法都是基于全参考的,而最切合实际的还要属无参考型图像质量评价。原始图像的获取困难和无参考模型设计困难直接推动了半参考型图像质量评价的发展。目前国内外研究最多的还是全参考型

图像质量评价,半参考型和无参考型虽然取得了许多成果,但是仍然不够完善,今后半参考型和无参考型图像质量评价将会成为研究热点,而无参考型图像质量评价将会成为最终的研究方向。

b) 图像质量评价体现的是人的主观感受,人类视觉系统是人们对图像的生理和心理方面的反应。评价图像质量的主体的主体是人,自然要考虑到人类是如何观察和感知图像的。因此人类视觉系统特性的研究将会是今后发展的重要方向。虽然现在对 HVS 特性的研究还很有限,而且其复杂程度难以模拟,但是它仍然是质量评价中最重要的量化模型,若能采用更加精确的 HVS 模型,评价性能应能进一步提高。虽然人们已经研究了人眼的一部分物理特性和心理特性并且很好地运用到了实际应用中,但是到目前为止研究得到的成果还处于低级阶段,人类视觉系统的研究还有漫漫长路。

c) 目前的全参考评价方法中的大部分都是针对彩色图像的亮度分量进行计算的,很少有评价方法在彩色空间上进行评价。但是现实世界是丰富多彩的,有时色彩对图像的视觉效果的影响很大,如何应用图像的色彩特征进行评价也是以后研究的方向。

d) 现有的图像质量评价的研究,都是在经历单一类型失真的图像上进行的,比如 LIVE 和 TID2008 中的退化图像都是假设只受到一种特定类型的退化,而实际上图像往往会经历多种失真类型。如何对多种失真类型的图像进行评价,并且要保持评价结果的稳定性,是更为复杂的研究课题。

e) 现有的全参考图像质量评价方法都是研究图像退化程度的,而没有研究图像增强程度的评价方法。全参考评价方法都是假设有一幅完美的标准图像,标准图像的质量要优于退化图像,通过把退化图像和完美图像进行比较来判定退化图像的退化程度。现实中经常需要对一幅效果不太理想的图像进行改善以增强视觉效果,往往需要某种评价方法用于度量各种增强方法。这时的参考图像效果要差于改善后的图像,就不能简单地应用现在研究出的各种方法,以后需要研究出参考图像并

非完美图像的方法。

f) 目前为止,国内外的学者们研究出了许多图像质量的评价方法,但是大多都是基于理论研究,没有一个权威并且可以通用到所有行业中的软件。任何研究都是以实际应用为目的的,所以实用性将会是将来发展的一个重要方向。为了把理论研究结果应用于实际应用中,就不能在对评价方法性能进行判定时,局限于准确性、单调性和一致性几个方面,还应当将方法的复杂度和运算效率及稳定性等因素考虑在内。有些方法虽然效果较好,但是计算时间较长,如何简化这些方法,使得计算机能够快速进行评价也需要进一步研究。

g) 在实际应用中,有时很难得到参考图像,因此,很多学者已经把研究方向转向半参考图像质量评价和无参考图像质量评价上。半参考图像质量评价只需要部分原始图像信息,无参考图像质量评价无需任何原始图像的信息,这大大降低了评价方法的时间复杂度和空间复杂度,并解决了获取原始图像困难的麻烦,但解决起来也更加困难。目前半参考和无参考图像质量评价方法还处在初步阶段,而且大部分评价方法要么只针对单一类型的失真进行,要么需要先进行某种形式的训练,评价效果达不到全参考评价方法所能达到的程度,训练模型和训练数据的选择也是一个问题。如何把全参考评价方法应用于无参考评价中,构造出一个用于比较的标准图像来使用全参考评价方法也是以后研究的一个方向。

h) 随着图像处理技术的发展,立体图像技术能提供深度信息来增强图像的立体感、临场感,是下一代数字媒体技术的发展方向。如何有效地对三维图像进行评价将是未来的研究热点和难点。

图像质量评价是图像处理领域的一个重要的研究方向,随着评价方法的逐渐成熟,图像质量评价方法的研究将会有广泛的应用前景,对图像和视频处理产生深刻的影响。本文对目前比较成熟的全参考图像质量评价方法进行了综述,对全参考评价方法的深刻理解对研究无参考及半参考评价方法有着重要的意义。

参考文献:

- [1] WANG Zhou, BOVIK A C. Modern image quality assessment [M]. New York: Morgan and Claypool Publishing Company, 2006: 20-30.
- [2] WANG Zhou, BOVIK A C, SHEIKH H R, *et al.* Image quality assessment: from error visibility to structural similarity [J]. IEEE Trans on Image Processing, 2004, 13(4): 600-612.
- [3] AVCIBAS I, SANBUR B, SAYOOD K. Statistical evaluation of image quality measures [J]. Journal of Electronic Imaging, 2002, 11(2): 206-213.
- [4] GIROD B. What's wrong with mean-squared error[M]//Digital Images and Human Vision. Cambridge: MIT Press, 1993: 207-220.
- [5] ESKICIOGLU A M, FISHER P S. Image quality measures and their performance [J]. IEEE Trans on Communications, 1995, 43(12): 2959-2965.
- [6] 庞全,王振华,耿丽硕,等.基于分形维数的图像质量客观评价方法研究[J].中国图象图形学报,2009,14(4): 657-662.
- [7] 庞建新,张荣,张晖,等.图像结构失真的编码质量评价[J].中国图象图形学报,2009,14(8): 1560-1568.
- [8] LI Bei, MEYER W, KLASSEN V R. A comparison of two image quality models[C]//Human Vision and Electronic Imaging. San Jose: SPIE Digital Library, 1998: 98-109.

- [9] DAMERA-VENKATA N, KITE T D, GEISLER W S, *et al.* Image quality assessment based on a degradation model [J]. IEEE Trans on Image Processing, 2000, 9(4): 636-650.
- [10] CHANDLER D M, HEMAMI S S. VSNR: a wavelet-based visual signal-to-noise ratio for natural images [J]. IEEE Trans on Image Process, 2007, 16(9): 2284-2298.
- [11] WANG Zhou, BOVIK A C. A universal image quality index [J]. IEEE Trans Signal Precessing, 2002, 9(3): 91-94.
- [12] WANG Zhou, SIMONCELLI P, BOVIC A C. Multiscale structural similarity for image quality assessment [C]//Proc of the 37th Asilomar Conference on Signals, Systems and Computers. 2003: 1398-1402.
- [13] LI Chao-feng, BOVIC A C. Three-component weighted structural similarity index [C]//Proc of SPIE. 2009: 72420Q-72420Q-9.
- [14] CHEN Guan-hao, YANG Chun-ling, XIE Sheng-li. Gradient-based structure similarity for image quality assessment [C]//Proc of International Conference on Image Processing. 2006: 2929-2932.
- [15] 朱里,李乔亮,张婷,等.基于结构相似性的图像质量评价方法[J].光电工程,2007,34(11): 108-113.
- [16] 王涛,高新波,张都应.一种基于内容的图像质量评价测度[J].中国图象图形学报,2007,12(6): 1002-1007.
- [17] 黄小乔,石俊生,杨健,等.基于色差的均方误差与峰值信噪比评价彩色图像质量研究[J].光子学报,2007,36(B06): 295-298.
- [18] 任雪,孙涵,张金国.一种新的基于局部特征的图像质量评价方法[J].中国图象图形学报,2010,15(8): 1236-1243.
- [19] YANG Wan, WU Le-hua, FAN Ye, *et al.* A method of image quality assessment based on region of interest [C]//Proc of the 7th World Congress on Intelligent Control and Automation. 2008: 6840-6843.
- [20] AJA-FERNANDEZ S, ESTEPAR R S, ALBEROLA-LOPEZ C, *et al.* Image quality assessment based on local variance [C]//Proc of the 28th Annual International Conference on Engineering in Medicine and Biology Society. Berlin: Springer-Verlag, 2006: 4815-4818.
- [21] 孔繁锦.结合HVS和相似性度量的图像质量评价测度[J].中国图象图形学报,2011,16(7): 1184-1191.
- [22] 叶盛楠,苏开娜,肖创柏,等.基于结构信息提取的图像质量评价[J].电子学报,2008,36(5): 856-861.
- [23] SHEIKH H R, BOVIC A C, De VECIANA G. An information fidelity criterion for image quality assessment using natural scene statistics [J]. IEEE Trans on Image Processing, 2005, 14(12): 2117-2128.
- [24] SHEIKH H R, BOVIC A C. Image information and visual quality [J]. IEEE Trans on Image Processing, 2006, 15(2): 430-444.
- [25] LIU Zheng, LAGANIRE R. Phase congruence measurement for image similarity assessment [J]. Pattern Recognition, 2007, 28(1): 166-172.
- [26] MORRONE M C, ROSS J, BURR D C, *et al.* Mach bands are phase dependent [J]. Nature, 1986, 324(6049): 250-253.
- [27] ZHANG Lin, ZHANG Lei, MOU Xuan-qin, *et al.* FSIM: a feature similarity index for image quality assessment [J]. IEEE Trans on Image Processing, 2011, 20(8): 2378-2386.
- [28] GAO Xin-bo, LU Wen, TAO Da-cheng, *et al.* Image quality assessment based on multiscale geometric analysis [J]. IEEE Trans on Image Processing, 2009, 18(7): 1409-1423.
- [29] LU Wen, GAO Xin-bo, LI Xue-long, *et al.* An image quality assessment metric based on contourlet [C]//Proc of the 15th IEEE International Conference on Image Processing. 2009: 1172-1175.

- [30] 孙玉宝,费选,韦志辉,等. 基于 Contourlet 的图像感知质量评价[J]. 电子学报, 2011, 39(3): 649-655.
- [31] LI Xue-long, LU Wen, TAO Da-cheng, *et al.* An image quality assessment based on wavelet structure [J]. *IEEE Trans on Signal Processing*, 2009, 89(4): 548-555.
- [32] CARNEC M, CALLET P L, BARBA D. Objective quality assessment of color images based on a generic perceptual reduced reference [J]. *Image Communication*, 2008, 23(4): 239-256.
- [33] LI Qiang, WANG Zhou. Reduced-reference image quality assessment using divisive normalization-based image representation [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2009, 3(2): 202-211.
- [34] GUNAWAN I P, GHANBARI M. Image quality assessment based on harmonics gain/loss information [C]//Proc of International Conference on Image Processing. Washington DC: IEEE Computer Society, 2005: 429-432.
- [35] WANG Zhou, SIMONCELLI E P. reduced-reference image quality assessment using a wavelet-domain natural image statistic model [C]//Human Vision and Electronic Imaging. San Jose: SPIE Digital Library, 2005: 149-159.
- [36] 张飞艳, 谢伟, 林立宇, 等. 基于小波域自然影像统计特性的无参考遥感影像质量评价[J]. 电子与信息学报, 2011, 33(11): 2742-2747.
- [37] WANG Zhou, WU Gui-xing, SHEIK H R, *et al.* Quality-aware images [J]. *IEEE Trans on Image Processings*, 2006, 15(6): 1680-1689.
- [38] WANG Sha, ZHENG Dong, ZHAO Ji-ying, *et al.* An image quality evaluation method based on digital watermarking [J]. *IEEE Trans on Circuits and Systems for Video Technology*, 2007, 17(1): 98-105.
- [39] WANG Zhou, BOVIK A C, EVANS B L. Blind measurement of blocking artifacts in images [C]//Proc of International Conference on Image Processing. 2000: 981-984.
- [40] LIU Shi-zhong, BOVIK A C. Efficient DCT-domain blind measurement and reduction of blocking artifacts [J]. *IEEE Trans on Circuits and Systems for Video Technology*, 2002, 12(12): 1139-1149.
- [41] WANG Zhou, SHEIKH H R, BOVIK A C. No-reference perceptual quality assessment of JPEG compressed images [C]//Proc of International Conference on Image Processing. 2002: 477-480.
- [42] MEESTERS L, MARTENS J B. A single-ended blockiness measure for JPEG-coded images [J]. *Signal Processing*, 2012, 82(3): 369-387.
- [43] MARZILIANO P, DUFAUX F, WINKLER S, *et al.* Perceptual blur and ringing metrics: applications to JPEG2000 [J]. *Signal Processing: Image Communication*, 2004, 19(2): 163-172.
- [44] SHEIKH H R, BOVIK A C, CORMACK L. No-reference quality assessment using natural scene statistics: JPEG2000 [J]. *IEEE Trans on Image Processing*, 2005, 14(11): 1918-1927.
- [45] 郭迎春, 于明. 基于子带相似性分析的 JPEG2000 图像无参考质量评价[J]. 电子与信息学报, 2011, 33(6): 1496-1500.
- [46] SAAD M A, BOVIK A C, CHARRIER C. A DCT statistics-based blind image quality index [J]. *IEEE Signal Processing Letters*, 2010, 17(6): 583-586.
- [47] SAAD M A, BOVIK A C, CHARRIER C. DCT statistics model-based blind image quality assessment [C]//Proc of the 18th IEEE International Conference on Image Processing. 2011: 3093-3096.
- [48] SAAD M A, BOVIK A C, CHARRIER C. Natural DCT statistics approach to no-reference image quality assessment [C]//Proc of 17th IEEE International Conference on Image Processing. 2010: 313-316.
- [49] MOORTHY A K, BOVIK A C. A two-step framework for constructing blind image quality indices [J]. *IEEE Trans on Signal Processing*, 2010, 17(5): 513-516.
- [50] MOORTHY A K, BOVIK A C. Blind image quality assessment: from natural scene statistics to perceptual quality [J]. *IEEE Trans on Image Processing*, 2011, 20(12): 3350-3364.
- [51] 楼斌, 沈海斌, 赵武锋, 等. 基于自然图像统计的无参考图像质量评价[J]. 浙江大学学报: 工学版, 2010, 4(2): 248-252.
- [52] 金波, 李朝锋, 吴小俊. 结合 NSS 和小波变换的无参考图像质量评价[J]. 中国图象图形学报, 2012, 17(1): 33-39.
- [53] 蒋刚毅, 黄大江, 王旭. 图像质量评价方法研究进展[J]. 电子与信息学报, 2010, 32(1): 219-226.
- [54] WAINWRIGHT M J, SIMONCELLI E P, WILSKY A S. Random cascades on wavelet trees and their use in analyzing and modeling natural images [J]. *Applied and Computational Harmonic Analysis*, 2001, 11(1): 89-123.
- [55] MORRONE M C, OWENS R A. Feature detection from local energy [J]. *Pattern Recognition*, 1987, 6(5): 303-313.
- [56] JAIN R, KASTURI R, SCHUNCK B G. *Machine vision* [M]. [S. l.]: McGraw-Hill Science/Engineering, 1995.
- [57] JAHNE B, HAUSSECKER H, GEIBLER P. *Handbook of computer vision and applications* [M]. Salt Lakecity: Academic Press, 1999.
- [58] YANG C, KWOK S H. Efficient gamut clipping for color image processing using LHS and YIQ [J]. *Optical Engineering*, 2003, 42(3): 701-711.
- [59] PONOMARENKO N, LUKIN V, ZELEENSKY A, *et al.* TID2008: a database for evaluation of full-reference visual quality assessment metrics [EB/OL]. [2009]. <http://www.ponomarenko.info/papers/mre2009tid.pdf>.
- [60] NINASSI A, CALLET P L, AUTRUSSEAU F. Categorical subjective image quality (CSIQ) database [DB/OL]. (2010-03). <http://vision.okstat.edu/?loc=csiq>.
- [61] SHEIKH H R, SESHADRINATHAN K, MOORTHY A K, *et al.* Image and video quality assessment research at LIVE [DB/OL]. [2004]. <http://live.ece.utexas.edu/research/quality/>.
- [62] NINASSI A, CALLET P L, AUTRUSSEAU. Subjective quality assessment: IVC database [DB/OL]. (2005) [2006]. <http://www.ircnyn.ec-nantes.fr/ivcd/>.
- [63] HORITA Y, SHIBATA K, KAWAYOKE Y, *et al.* MICT image quality evaluation database [DB/OL]. [2000]. <http://mict.eng.u-toyama.ac.jp/mict/index2.html>.
- [64] CHANDLER D M, HEMAMI S S. A57 database [DB/OL]. [2007]. <http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html>.
- [65] VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment [EB/OL]. (2003-03). <ftp://vqeg.its.bldrdoc.gov/Documents/Meetings/Hill-sboro-VQEG-Mar-03/VQEGII-DraftReportv2a.pdf>.
- [66] SHEIKH, SABIR M F, BOVIK A C. A statistical evaluation of recent full reference image quality assessment algorithms [J]. *IEEE Trans on Image Processing*, 2006, 15(11): 3440-3451.