

无参考图像质量评价

朱文斌, 陈 强, 杨曦晨

(南京理工大学 计算机科学与工程学院, 江苏 南京 210094)

摘要: 图像质量评价是近几年图像处理领域比较热门的研究课题。目前,许多学者已经提出了各种各样的无参考质量评价方法。对无参考方法进行综述,详细介绍 BIQI, BLIINDS-II, BRISQUE, DESIQUE, DIVINE, NIQE, SSEQ 等无参考质量评价方法,并在 LIVE 和 TID2008 数据库上进行实验分析,最后根据分析的结果探讨图像质量评价的发展方向。

关键词: 图像质量评价; 无参考; 自然场景统计特征; 变换域

中图分类号: TN911.73-34; TP391

文献标识码: A

文章编号: 1004-373X(2015)18-0081-08

No-reference assessment algorithms for image quality

ZHU Wen-bin, CHEN Qiang, YANG Xi-chen

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

Abstract: Image quality assessment is a hot research topic in the field of image processing in recent years. Many scholars have proposed a variety of non-reference image quality evaluation methods. Some non-reference image quality assessment algorithms are summarized in this paper, such as BIQI, BLIINDS-II, BRISQUE, DESIQUE, DIVINE, NIQE and SSEQ. They were tested and analyzed on LIVE and TID2008 database. The development direction for image quality assessment is discussed in this paper according to the analysis results.

Keywords: image quality assessment; no-reference; natural scene statistic; transform domain

0 引言

数字图像在获取、存储、传输、显示的过程中,都有可能引入失真,导致图像质量的下降。失真图像丢失了一些重要信息,可能导致满足不了应用与系统要求^[1]。特别是随着互联网的快速发展,在网络中传输的图像因带宽、时效等限制条件,图像经常会受到压缩比例较大的压缩而造成严重的失真。因此,如何准确地衡量图像质量的好坏,成为了计算机视觉以及图像处理领域热门且具有挑战性的问题。

按照是否需要人的参与,图像质量评价方法可以分为两类:主观图像质量评价和客观图像质量评价^[2]。主观图像质量评价由人眼视觉效果对图像进行评价;与此对应,客观图像质量评价算法则基于特定算法由计算设备对图像进行自动评价。主观质量评价一般由大量的观察者对图像打分,且易受评价环境、评价参与者知识

水平以及评价人员情绪的影响。主观评价结果随机性较强,很难在实时图像系统中应用^[3]。基于主观质量评价的局限性,客观质量评价算法被广泛研究。客观质量评价方法不需要人工参与,因此能被用到各种图像处理算法的性能评价和实时图像系统中。研究高效、优秀的图像质量评价算法,具有很高的现实意义。

根据能够得到的参考图像(原始图像)信息的多少,客观的图像质量评价方法分为全参考质量评价方法、半参考质量评价方法和无参考质量评价方法^[4]。全参考质量评价方法,指的是能获得参考图像所有信息的评价方法。通过计算待评价的图像与原始图像之间的差异性,对图像进行评价。全参考方法目前已经比较成熟可靠,半参考评价方法能获得原始图像的部分信息,无参考评价方法,则无法获得参考图像的任何信息。本文主要讨论无参考评价方法。

1 相关模型介绍

1.1 图像对比度归一化^[5]

对于一幅灰度图像 $I(M \times N)$, 对其进行去均值与归一化(如式(1)所示),得到其 MSCN (Mean Subtracted Contrast Normalized), 即 \hat{I} 。

收稿日期: 2015-03-25

基金项目: 十二五民用航天技术预先研究项目
(D040201); 中国航天科技集团公司航天科技
创新基金资助项目(casc05131418)

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \quad (1)$$

式中: $i \in \{1, 2, \dots, M\}, j \in \{1, 2, \dots, N\}$, M, N 分别是图像的高和宽。 C 是常数 1, 防止分母为 0。变量 $\mu(i, j)$ 和 $\sigma(i, j)$ 定义如下:

$$\mu(i,j) = \sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} I(i+k, j+l) \quad (2)$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} [I(i+k, j+l) - \mu(i,j)]^2} \quad (3)$$

$$w = \{w_{k,l} | k = -K, \dots, K, l = -L, \dots, L\} \quad (4)$$

式中 w 是一个 2 维的高斯权重函数。通常情况下 K, L 取 3。如图 1 所示。

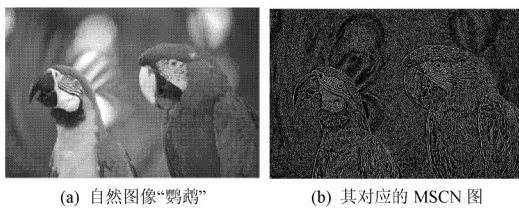


图 1 自然图像及其对应的 MSCN

1.2 对称广义高斯模型和均值非零的非对称广义高斯分布模型

自然图像统计特征直方图满足高斯分布, 因此高斯模型可以被用来模型化实验数据。高斯模型分为均值为零的对称广义高斯模型 (Generalized Gaussian Distribution, GGD) 和均值非零的非对称广义高斯分布模型 (Asymmetric Generalized Gaussian Distribution, AGGD)。

如图 2 所示, 图 1(b) 的 MSCN 的统计直方图满足高斯分布, 其可以用 GGD 模型进行模拟。

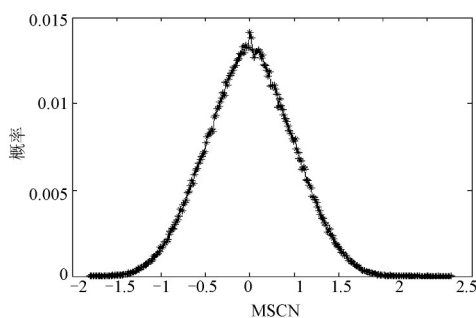


图 2 MSCN 的统计直方图

GGD 模型如式 (5) 所示, σ^2 是分布的方差, α 控制着分布的形状, 例如当 $\alpha=2$ 时, 此分布是高斯分布, $\alpha=1$ 时此分布是拉普拉斯分布。GGD(α, σ^2) 参数对可以使用文献 [6] 中的瞬间匹配方法快速计算得到。

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^\alpha\right) \quad (5)$$

$$\beta = \sigma \sqrt{\Gamma(1/\alpha)/\Gamma(3/\alpha)} \quad (6)$$

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad x > 0 \quad (7)$$

式中: $\Gamma(\cdot)$ 是伽马函数; β 是与 σ 相关的变量; t 是任意一个取值 $[0, +\infty]$ 的变量。

AGGD 模型如式 (8) 所示, γ 控制着分布的形状, σ_l^2, σ_r^2 是尺度参数, 分别控制着模型两侧分布。 η 是分布的均值。如果 $\sigma_l^2 = \sigma_r^2$, 那么 AGGD 就成为了 GGD, 可知此时均值 η 为零。

$$f(x; \gamma, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{\gamma}{(\beta_l + \beta_r)\Gamma(\frac{1}{\gamma})} \exp\left(-\left(\frac{-x}{\beta_l}\right)^\gamma\right), & \forall x \leq 0 \\ \frac{\gamma}{(\beta_l + \beta_r)\Gamma(\frac{1}{\gamma})} \exp\left(-\left(\frac{x}{\beta_r}\right)^\gamma\right), & \forall x \geq 0 \end{cases} \quad (8)$$

其中:

$$\beta_l = \sigma_l \sqrt{\Gamma(1/\gamma)/\Gamma(3/\gamma)}, \quad \beta_r = \sigma_r \sqrt{\Gamma(1/\gamma)/\Gamma(3/\gamma)} \quad (9)$$

$$\eta = (\beta_r - \beta_l)\Gamma(2/\gamma)/\Gamma(1/\gamma) \quad (10)$$

2 方法介绍

2.1 NIQE^[7]

2.1.1 特征介绍

假设某一幅输入图像 $I(M \times N)$ 。该方法首先对图像进行去均值与标准化, 得到图像的 MSCN 即 \hat{I} 。研究发现, 质量较好的自然图像的统计特征满足高斯模型, 然而失真会破坏这种模型, 因而这种变化的程度会体现在人的感知上。如图 3 所示, 5 种失真图像的统计直方图都与参考图像有着较大的差异。

NIQE 方法采用 96×96 块间无重合的方式对 \hat{I} 进行分块。由于人眼对于对比度比较大的图像区域更加的敏感, 因此对前述得到的这些图像块进行筛选。计算每个块的局部对比度均值 $\sigma(b) = \sum_{(i,j) \in \text{patchb}} \sigma(i,j)$, 筛选出 $\sigma(b) > T$ ($T = p \cdot \max(\sigma(i,j)), i \in [1, M]; j \in [1, N]; p = 0.75$) 的图像块进行特征提取。对上述步骤筛选得到的图像块, 使用对称广义高斯模型 (GGD, 式 (5)) 进行建模, 得到基于空域点的特征 GGD(α, β)。同时分别从横向、纵向、主对角线、副对角线 4 个方向对相邻像素使用非对称广义高斯分布模型进行 (AGGD, 式 (8)) 进行建模, 分别得到特征 AGGD($\gamma, \beta_l, \beta_r, \eta$), 共得到 16 个参数。上述步骤一共得到 18 个特征。对同一个图像使用最近邻算法进行下采样, 用同样的方法得到 18 个参数, 一幅图像共得到 36 个特征。

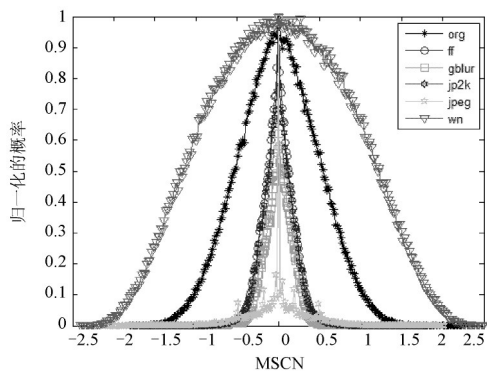


图3 五种失真图像的 MSCN 统计直方图

2.1.2 模型训练与评价图像

本方法对训练集中的原始高质量自然图像进行特征提取,获得一个 $N \times 36$ 的特征矩阵。对这个矩阵求均值得到一个 1×36 行向量 \mathbf{v}_1 以及求协方差得到协方差矩阵 Σ_1 。作为模型用于评价图像质量。对于待评价的图像,同样进行上述特征提取的所有步骤(除了没有块筛选步骤),得到一个行向量 \mathbf{v}_2 以及协方差矩阵 Σ_2 。

通过计算理想模型与失真图像参数之间的差异,如式(11),得到最终的质量评价分数:

$$\text{Index} = \sqrt{(\mathbf{v}_1 - \mathbf{v}_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (\mathbf{v}_1 - \mathbf{v}_2)} \quad (11)$$

式中 $\mathbf{v}_1, \mathbf{v}_2$ 和 Σ_1, Σ_2 分别为理想模型和实际图像的均值向量与协方差矩阵。

2.2 BRISQUE^[8]

2.2.1 特征介绍

此方法类似于前述的 NIQE 方法,但是没有分块步骤。对于训练图像 I ,首先计算其 MSCN 得到 \hat{I} ,然后对 \hat{I} 进行空域广义高斯模型 GGD 建模得到参数 $\text{GGD}(\alpha, \sigma^2)$ 。接着再在横向、纵向、主对角线、副对角线 4 个方向上使用非对称广义高斯分布模型 AGGD 进行建模,分别得到 $\text{AGGD}(\gamma, \sigma_l^2, \sigma_r^2, \eta)$,共得到 16 个特征。然后使用最近邻算法进行下采样,在另一个尺度上对提取特征,最终训练图像获得 36 的特征。

2.2.2 模型训练与分值计算

将图像数据集分成训练集和测试集,将训练集的图像利用上述的方法进行特征提取,得到特征向量。利用 Libsvm 对训练集中图像的特征向量及其对应 MOS 值进行支持向量回归学习^[3,9],得到图像评价模型。对于某一个测试图像,按照前述方法提取出特征向量,再使用得到的图像质量评价模型,回归计算得到最终的分值 Index。

2.3 DESIQUE^[9]

2.3.1 特征介绍

本方法分别基于频域与空域进行特征提取:

基于空域的特征提取:首先计算图像的 MSCN,得到 \hat{I} 。然后对其进行广义高斯模型 GGD 拟合,得到特征参数 $\text{GGD}(\alpha, \sigma^2)$ 。然后对 \hat{I} 进行对数化得到 $J(i, j)$ 如式(12)所示:

$$J(i, j) = \ln(|\hat{I}(i, j)| + K), \quad K = 0.1 \quad (12)$$

分别在 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 以及主对角线减副对角线 5 个方向上对 $J(i, j)$ 进行 GGD 拟合,分别得到 $\text{GGD}(\alpha, \sigma^2)$,共获得 10 个特征参数。

基于频域的特征提取:在频域,首先对图像在 2 个方向(水平和垂直)方向上进行了 log-Gabor 变换,得到 2 个 log-Gabor 子带 $g(i, j)$ 。然后对这 2 个子带分别进行对数化得到 $J(i, j)$,如式(13)所示:

$$J(i, j) = \ln(|g(i, j)| + K), \quad K = 0.1 \quad (13)$$

其中 $|g(i, j)|$ 代表的是模。最后对 2 个子带 $J(i, j)$,分别在 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 以及主对角线减副对角线 5 个方向上对 $J(i, j)$ 进行 GGD 拟合,分别得到 $\text{GGD}(\alpha, \sigma^2)$,共获得 20 个参数(2 个参数 \times 5 个方向 \times 2 个子带)。

最后对图像使用最近邻算法进行下采样,提取出 20 个特征参数,DESIQUE 方法一共得到 60 个特征参数。

2.3.2 模型训练与分值计算

DESIQUE 将图像数据库分为训练集和测试集 2 部分。方法首先对图像进行特征提取,得到训练图像特征集。模型训练分 2 步进行,分别为不针对特定失真类型的训练和针对特定失真类型的训练。第 1 步是将所有训练图的特征及其对应的 DMOS 利用 SVM 进行分类训练,得到 Model-Svm。假设图像数据库有 N 种失真类型,第 2 步是针对特定的失真,将图像根据其真实的失真类型分类,每一类失真类型的图像分别与其对应 DMOS 分别使用 SVR 进行回归,得到此失真类型的回归模型 Model- n ,其中 n 取 $1 \sim N$ 。

对测试图像分别进行频域与空域的特征提取。其评价由下步骤获得:首先,假设不知道测试图像的失真类型,然后使用 SVM 基于 model-svm 对图像进行分类,计算其属于各种失真类型的概率,记为向量 $\mathbf{P}(p_1, p_2, \dots, p_N)$ 。再分别假设其属于某个特定失真类型,基于 Model- n ,利用 SVR 模型,分别计算其分数,组成分数向量 $\mathbf{S}(s_1, s_2, \dots, s_N)$ 。最后,计算待评价图像的质量分数为 Index,如式(14)所示:

$$\text{Index} = \mathbf{P} \cdot \mathbf{S} = \sum_{i=1}^N p_i \times s_i \quad (14)$$

2.4 BIQI^[10]

2.4.1 特征介绍

首先使用 Daubechies 9/7 小波基对图像进行 3 个

尺度,3个方向的变换,得到9个子带。NSS研究表明,经过这种变换后,每个子带的系数呈现出拉普拉斯分布,并可以使用广义高斯分布GGD模型进行参数化。它有3个参数,分别是均值 μ 、形状参数 α 、方差 σ^2 ,因为小波变换后的系数均值为0,所以每个子带可以得到特征参数 α, σ^2 。因此一幅图像一共可以得到18个特征(3尺度 \times 3方向 \times 2参数)。

2.4.2 模型训练与分值计算

本方法使用了和DESIQUE相同的2步骤模型训练与计算分数方法得到测试图像的分值Index。

2.5 DIIVINE^[11]

2.5.1 特征介绍

本方法使用可转向金字塔小波分解对图像在2个尺度6个方向上进行分解,得到12个分解结果子带,记为 S_a^θ ,其中 a 代表尺度,可取1和2, θ 代表方向,可以取 $\{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$ 。然后对每一个子带进行正常化,得到 d_a^θ 。此方法利用 d_a^θ 一共提取88个图像特征用于质量评价。

(1) 对每一个 d_a^θ ,使用广义高斯分布模型GGD对其进行模拟,因为金字塔小波变换后的系数均值为0,因此每个子频带可以获得参数GGD(σ^2, γ),12个子频带共获得24个参数,组成 $f_1 \sim f_{24}$ 。

(2) 图像天然是多尺度的。不同尺度,同一个方向的子频带间存在着某种关系。不同尺度,同一个方向的子频带的系数,被结合到一起,同样进行广义高斯分布模型GGD拟合,6个方向共得到12参数。因为 σ^2 不会给分析增加比 $f_1 \sim f_{12}$ 多的任何信息,因此6个方向只留下 γ 作为特征,得到 $f_{25} \sim f_{30}$,而后最终将所有的子频带系数结合到一起,进行GGD拟合得到最后的一个 γ ,即 f_{31} 。

(3) 实验证明,图像的高通响应和带通响应之间有着一定的统计学关系,而图像的失真会破坏这种关系。因此每一个子频带可以计算得到如此一个系数 ρ , $\rho = (2\sigma_{xy} + C_2) / (\sigma_x^2 + \sigma_y^2 + C_2)$ 。对于每个子频带,使用 15×15 的高斯滤波器对其高通响应和带通响应进行滤波, $\sigma_x^2, \sigma_y^2, \sigma_{xy}$ 分别是高通响应局部方差、带通响应局部方差和两者之间的互协方差。每个子带的所有局部 ρ 的均值就是这个子带的特征,因此一共得到12个特征,即 $f_{32} \sim f_{43}$ 。

(4) 自然图像有着一种相关结构,随着距离函数的变化而缓慢变化。将尺度为1的各子带中某个点以及和它距离在一定范围内的点,计算联合分布,可以得到一个三项式。这个三项式的3个系数以及参数间的错误值,多项式的真实值构成了一个子频带的参数,因此一共可以获得30个参数,即 $f_{44} \sim f_{73}$ 。

(5) 计算同尺度,不同方向间的子频带之间的关系,得到 $C_6^5 = 15$ 个特征,即 $f_{74} \sim f_{88}$ 。

2.5.2 模型训练与分值计算

本方法使用了和DESIQUE、BIQI相同的两步骤模型训练与计算分数方法得到测试图像的分值Index。

2.6 SSEQ^[12]

SSEQ方法利用了图像块间的统计学规律,通过计算图像块的信息熵,获得图像中的结构以及图像中像素点相互关系的大小。在进行大量统计与训练的情况下,对测试图像进行质量预测。

2.6.1 特征介绍

SSEQ在3个尺度上一共要提取12个特征。对于输入图像,首先对图像进行分块,计算每一块的区域的局部熵,获得所有局部熵值的平均数,作为特征1;然后对这些局部熵值排序,求出其偏度,作为特征2;再对同样分块的图像,计算每个局部块的局部谱域熵,获得所有局部谱域熵的平均数,作为特征3;然后对这些局部谱域熵排序,求出其偏度,作为特征4。使用最近邻算法对图像进行2次下采样,分别计算得到类似的4个特征,一共获得12个图像的特征。

2.6.2 模型训练与图像质量分计算

SSEQ使用了和DESIQUE、BIQI、DIVINE相同的方法得到测试图像的分值Index。

2.7 BLIINDS-II^[13]

2.7.1 特征介绍

本方法一共提取4个特征用于计算质量分数。

(1) 对图像进行大小为 5×5 ,块间重合为2的分块,对每一个图像块进行DCT变换得到相应的DCT系数。

(2) 对DCT系数进行GGD拟合,获得每个图像块的形状参数 γ ,取其中最低的10% γ 值的均值作为第1个特征,所有 γ 的均值作为第2个特征;提出了一个新的参数 ζ ,其计算式为:

$$\zeta = \frac{\sigma|X|}{\mu|X|} = \sqrt{\frac{\Gamma(1/\gamma)\Gamma(3/\lambda)}{\Gamma^2(2/\gamma)}} - 1 \quad (15)$$

其中 X 为每一个DCT块的系数, $\sigma|X|$ 和 $\mu|X|$ 分别是该块的方差和均值, $\Gamma(\cdot)$ 是前述的伽马函数。对每一个图像块,计算其对应的 ζ ,取其中最低的10% ζ 值的均值作为第3个特征,所有 ζ 的均值作为第4个特征。

(3) 将每个图像块中的DCT系数按照径向空间频率分成低、中、高3个部分,对3个部分的值分别求解方差,得到 σ_n^2 ,设 $E_n = \sigma_n^2$,低、中、高频率分别代表1,2,3。设 $R_n = \left| E_n - \frac{1}{n-1} \sum_{j < n} E_j \right| / \left(E_n + \frac{1}{n-1} \sum_{j < n} E_j \right)$,取较大

的 10%的 R_n 的均值作为第 5 个特征,所有 R_n 的均值作为第 6 个特征。

(4) 按照和径向空间频率变化垂直的方向,将每一个图像块获得的 DCT 系数分为 3 个部分,分别计算 3 个部分的 ζ ,全部图像块的所有 ζ 的最低的 10%的均值作为第 7 个特征,所有 ζ 的均值作为第 8 个特征。

2.7.2 模型训练与图像质量分计算

使用简单的概率模型,设 X 为特征向量,计算 X 与 DMOS 的概率分布,由于 $P(X,DMOS)=P(DMOS|X)P(X)$,计算获得的 $P(DMOS|X)$ 即为质量分数。使用数据集的 80%进行训练,20%的数据进行测试。

3 实验结果与分析

3.1 实验数据库

本文在 LIVE 和 TID 2008 两个数据库上对 BIQI^[10], NIQE^[7], BLIINDS - II^[13], BRISQUE^[8], DESIQUE^[9], DIIVINE^[11],SSEQ^[12]7 种方法进行测试。

LIVE 数据库^[14]包含 29 张参考图像以及对应的 5 种失真类型的共 779 张失真图像。5 种失真类型分别为 JPEG2000 压缩失真、JPEG 压缩失真、白噪声 (White_Noise, WN)、高斯模糊 (Gaussian_Blur, GBLUR)、快衰弱失真 (Fast Fading, FF)。LIVE 提供了图像主观评价分值 MOS。MOS 值取值范围为[0, 100]。一般实验采用 100 减去 MOS 值计算得到的 DMOS 进行算法测试。

TID 2008 数据库^[15]包含了 25 张参考图像以及 17 种失真类型的 1 700 张失真图像。其中,每一张参考图像对应 68 张失真图像,17 种失真,每一种失真有 4 个失真级别。

TID2008 和 LIVE 数据库都拥有 JPEG2000 压缩失真、JPEG 压缩失真、白噪声和高斯模糊 4 种失真类型的失真图像,LIVE 独有快衰弱失真。因此,为了方便与 LIVE 数据库进行对比,本文只针对 2 个数据库共同拥有的失真类型进行实验分析。

3.2 评价标准

本文使用斯皮尔曼等级次序相关系数 (SPROCC) 和皮尔逊线性相关系数 (LCC) 来对评价算法评价结果与数据库中主观分值的一致性进行衡量。主观评分与客观评分线性相关度越好,算法的表现越好。

3.3 实验结果

本文主要在 LIVE 数据库上进行模型训练,并使用训练得到的模型分别在 LIVE 数据库和 TID 数据库上进行对比实验。本文中介绍的方法都需要训练得到回归模型,为了算法间的平等比较,将 LIVE 数据库随机划分

为 80%的训练子集与 20%的测试子集,2 个集合没有任何的重合。该过程被重复了 1 000 次,以确保没有因为训练的空域内容导致的偏差。

表 1~表 4 分别表示上述 7 种无参考方法在 LIVE 数据库测试集和 TID2008 数据库上的运行结果。

表 1 7 种方法在 LIVE 数据库上的 LCC

	JP2k	JPEG	WN	GBLUR	FF	ALL
BIQI	0.872 6	0.830 1	0.927 7	0.929 6	0.776 4	0.867 3
BLIINDS- II	0.942 3	0.900 4	0.943 7	0.926 9	0.882 9	0.919 2
BRISQUE	0.933 7	0.882 3	0.964 3	0.959 3	0.907 5	0.929 4
DESIQUE	0.931 4	0.881 9	0.966 6	0.978 8	0.915 9	0.934 9
DIIVINE	0.827 7	0.738 4	0.959 8	0.956 1	0.839 0	0.864 2
NIQE	0.881 9	0.712 2	0.772 5	0.874 4	0.838 3	0.815 9
SSEQ	0.941 7	0.971 0	0.967 5	0.960 5	0.881 2	0.944 4

表 2 7 种方法在 LIVE 数据库上的 SPROCC

	JP2k	JPEG	WN	GBLUR	FF	ALL
BIQI	0.902 3	0.912 1	0.960 0	0.963 2	0.821 7	0.911 8
BLIINDS- II	0.942 0	0.907 6	0.970 2	0.934 5	0.895 7	0.930 0
BRISQUE	0.938 5	0.922 8	0.986 4	0.978 4	0.929 4	0.951 1
DESIQUE	0.947 2	0.916 9	0.992 8	0.994 6	0.949 8	0.960 3
DIIVINE	0.849 1	0.810 7	0.979 6	0.969 7	0.846 2	0.891 0
NIQE	0.906 3	0.846 0	0.974 8	0.944 9	0.882 1	0.910 8
SSEQ	0.941 4	0.958 4	0.981 6	0.953 5	0.895 6	0.946 1

3.3.1 算法表现比较

从结果可以看出,几种无参考评价方法虽然表现了较好的整体效果。但是对于不同的失真类型,评价结果存在较大差异。这主要与算法提取的特征有关。某些特征也许和某一个失真类型关系较大,而与另一个失真类型关系较小。

从表 1~表 4 可以看出,对于各失真类型,DESIQUE 方法都表现的较好,其在空域和频域分别进行了多方向多尺度的特征提取,对各种失真具有较大的敏感性。对于 JP2k, JPEG 失真, BLIINDS- II, DESIQUE, SSEQ 表现较好。BLIINDS- II 提取的特征较多的关注了参数模型的形状参数 γ , 而压缩导致的块效应会显著影响这个参数。压缩也导致了块内熵值显著变化,使得 SSEQ 方法表现优秀。对于 WN 失真, BRISQUE, DESIQUE 及 BIQI 表现较好。对于 GBLUR 失真, SSEQ 与 DESIQUE 方法表现较为出色。高斯模糊对图像熵的变化影响较大,使得以熵作为特征的 SSEQ 方法表现出色。

3.3.2 数据库依赖性

从表 1~表 4 中可以看出,上述各方法采用 LIVE 数

据集训练得到的模型,在TID2008数据库上的表现较差。例如BIQI方法在LIVE数据库中平均LCC为0.867 3,而使用LIVE训练的模型在TID2008数据库上测试,LCC只有0.781 9。

表3 7种方法在TID2008数据库上的LCC

	JP2k	JPEG	WN	GBLUR	ALL
BIQI	0.671 5	0.838 9	0.896 5	0.720 4	0.781 9
BLIINDS-II	0.896 5	0.890 9	0.858 3	0.806 7	0.863 1
BRISQUE	0.864 5	0.894 1	0.643 8	0.737 5	0.785 0
DESIQUE	0.920 9	0.946 6	0.871 2	0.841 9	0.895 1
DIIVINE	0.828 2	0.593 7	0.841 7	0.813 7	0.769 3
NIQE	0.879 1	0.817 6	0.842 0	0.816 5	0.838 8
SSEQ	0.873 8	0.839 5	0.834 8	0.844 6	0.848 2

表4 7种方法在TID2008数据库上的SPROCC

	JP2k	JPEG	WN	GBLUR	ALL
BIQI	0.694 0	0.844 3	0.897 1	0.746 8	0.795 5
BLIINDS-II	0.900 9	0.878 1	0.866 1	0.838 7	0.870 9
BRISQUE	0.901 6	0.875 3	0.623 4	0.798 9	0.799 8
DESIQUE	0.929 1	0.916 4	0.887 8	0.857 3	0.897 6
DIIVINE	0.852 5	0.630 9	0.872 1	0.823 7	0.794 8
NIQE	0.896 5	0.856 9	0.853 9	0.814 6	0.855 5
SSEQ	0.900 4	0.840 3	0.877 2	0.845 1	0.865 8

从上述表中可以看出,DESIQUE,SSEQ,BLIIND-II,NIQE具有较小的数据库依赖性,其在TID2008数据库中的LCC与SPROCC仍能保持在0.8以上,符合预期。而BIQI、BRISQUE与DIIVINE具有较大的数据库依赖性,其在TID2008数据库中的LCC与SPROCC出现了小于0.8的情况。

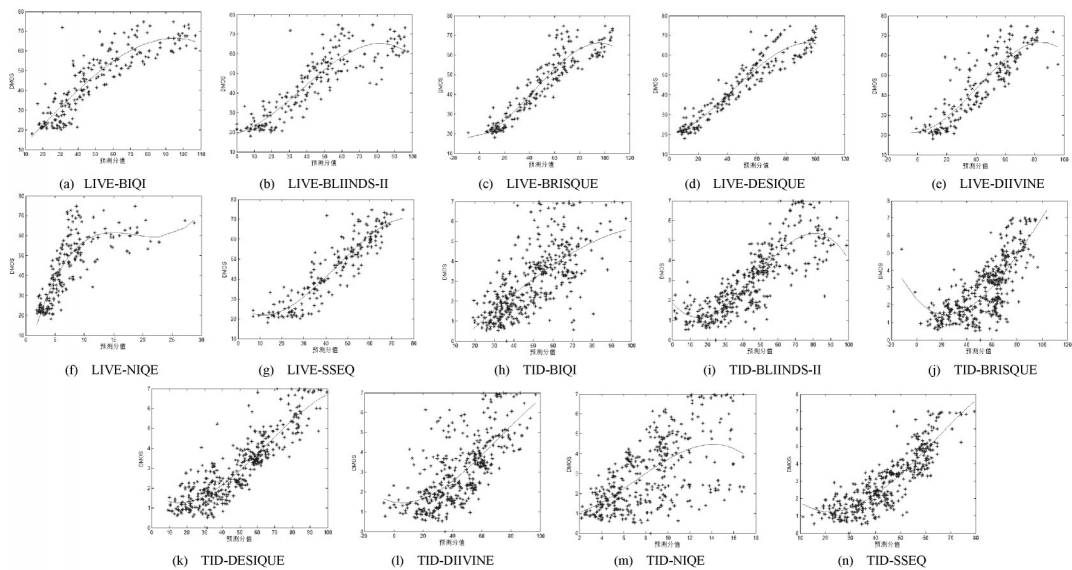


图4 7种方法在LIVE和TID2008数据库上评价与DMOS值的关系曲线

图4是7种方法在LIVE数据库以及TID2008数据库上的散点图。图中可以看出,BRISQUE,DESIQUE,SSEQ的结果与DMOS有着较好的线性关系,且具有较低离群度。而BIQI,DIIVINE,NIQE与BLIINDS-II的点离散度很大,说明算法结果与DMOS的关系不密切。

3.3.3 算法效率对比分析

本文对7种无参考方法分别在不同尺寸的图像上进行了实验,进行算法效率的比较。实验使用了4张尺寸分别为256×256,512×512,1 024×1 024,1 600×1 600的图像。7种算法分别在这4张图像上进行100次实验,运行的平均时间如表5所示。本文实验使用PentiumE2140@1.60 GHz,Windows 7 Pro 64-bit,3 GB内存的

计算机,运行软件为Matlab 2013a。

表5 7种方法的运行时间对比

	256×256	512×512	1 024×1 024	1 600×1 600
BIQI	2.365	2.599	3.044	5.210
BLIINDS-II	53.969	190.228	929.865	1 811.585
BRISQUE	0.466	0.609	1.383	2.744
DESIQUE	0.407	1.372	4.810	11.792
DIIVINE	26.383	60.176	207.883	559.822
NIQE	0.169	0.786	2.950	7.638
SSEQ	1.248	4.894	19.277	47.041

从表5中可以看出,上述7种无参考质量评价方法

中, DESIQUE, BRISQUE, NIQE 方法因为仅仅在空域和频域上进行直观的特征提取,没有太多的中间计算环节,因此具有较高的效率。BIQI 与 SSEQ 方法效率虽然略低于前述 3 个方法,但仍然在一个数量级。而 BLIINDS- II 和 DIIVINE 因为在频域上进行大量的特征提取以及进行许多中间参数计算,导致效率极低。特别地, BLIINDS- II 方法消耗时间甚至达到了前述 5 种方法的 100 倍以上。

图 5 表示了各种方法的运行效率随着图像尺寸的变化而变化的情况。其中,以各方法在 256×256 情况下运行的时间为单位 1。从表中可以看出 BIQI 方法的运行效率受图像尺寸的影响是最小的,运行时间增长不明显。而另外 6 种方法运行效率受图像尺寸的影响明显。结合表 5 和图 5 可以看出,虽然在 256×256 的尺寸下, NIQE 方法的运行效率最高,但是随着图像尺寸的增大, BRISQUE, BIQI 表现出了更高的效率。因此,在图像尺寸较小的情况下, BIQI, BRISQUE, DESIQUE, NIQE, SSEQ 的表现都让人满意,而在图像尺寸很大的情况下, BRISQUE, BIQI 方法有着明显的效率优势。

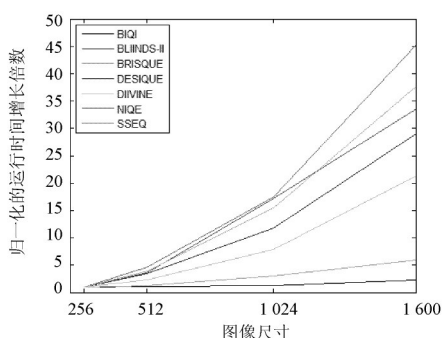


图 5 7 种方法运行时间随图像尺寸的变化

4 结 语

本文对无参考图像质量评价方法进行综述,对 BIQI, NIQE, BLIINDS - II, BRISQUE, DESIQUE, DIIVINE, SSEQ 等无参考质量评价方法进行了详细的描述,并在 LIVE 和 TID 2008 数据库上对其进行了测试,对各种方法进行详细分析以及比较,讨论了各方法的优缺点。

今后图像质量评价的发展有以下几个方面亟待解决:

(1) 目前优秀的无参考的质量评价方法很多都是基于自然图像统计 NSS 模型的,其中的关键是相关参数的广义高斯模型拟合,未来可以提出新的拟合模型,使得其更加符合统计数据分布,以达到提升现有的算法性能的目标。

(2) 建立完善的 HVS 模型:使用 HVS 模型分析视

觉输入的原理在图像质量评价中已经起到了举足轻重的地位。然而,目前关于 HVS 的理解以及采用的计算模型还处于非常低级的阶段。绝大多数现有的基于生物学的初级视觉皮层 (Primary Visual Cortex, V1) 的模型并不是很完善。未来应该在这方面进行努力,改进基于 V1 的模型,研究人眼的物理特性和心理特性之间的关系并运用到了实际应用中去。

(3) 研究对复合型的图像失真的评价方法。图像质量评价中一个很大的挑战在于图像失真往往是复合型的,同时其失真程度也是不可控的。尽管许多的算法已经被用来评价单种失真类型的图像,能够评价混合失真类型的算法显然更加实用。一个 IQA 不仅要考虑这些失真混合产生的联合作用,同时还要考虑这些失真互相之间的影响。在图像处理过程中会产生各种各样的图像失真,理想的 IQA 算法应该完全能够处理这些失真。同时,现在 IQA 面临的一个新问题是调整现有的模型来处理超阈值失真。

(4) 当评价一个图像质量时,根据失真的多少以及其是否影响了图像对象的显著外观, HVS 使用的可视化策略会有所不一样。许多研究表面 HVS 是一个高度自适应的系统。其能够基于失真与图像之间的相互作用等因素做出适当的调整。因此,研究图像与失真之间的关系,设计自适应视觉策略进行质量评价是一个很有研究意义的方面。

(5) 目前绝大多数 IQA 算法都无法处理几何失真,例如平移、缩放、旋转、剪切等。这些失真,假如不是很强烈,一般对图像的视觉质量影响很小。然而,即使是轻微的几何变化,会产生大规模的像素强度逐点变化,因而大多数的现有 IQA 算法对这种轻微几何失真的图像的质量评价要比图像的实际质量低很多。由于 IQA 已经被用到了越来越多的领域,对几何失真的处理显然变得越来越重要。

(6) 其实最具挑战性的是算法的运行效率。随着 IQA 从研究领域到实际应用领域进步,计算速度、内存占用等有关效率方面的因素成为评价一个算法优劣的重要标准之一。有些图像质量评价算法还会被应用于处理视频帧或者一些实时系统上,因此效率非常重要。

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,其会对图像和视频处理产生深远的积极影响。本文对目前已提出的无参考图像质量评价方法进行了综述,并对图像质量评价的未来发展及挑战进行了总结。对图像质量评价方法的深刻理解对于研究图像处理、模式识别等有着重要的意义。

参 考 文 献

- [1] 梁狄林.无参考图像质量评价算法研究[D].无锡:江南大学, 2013.
- [2] 褚江,陈强,杨曦晨.全参考图像质量评价综述[J].计算机应用研究,2014,31(1):13-22.
- [3] SHEIKH H R, SABIR M F, BOVIK A C. A statistical evaluation of recent full reference image quality assessment algorithms [J]. IEEE Transactions on Image Processing, 2006, 15(11): 3440-3451.
- [4] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [5] RUDERMAN D L. The statistics of natural images [J]. Network Computation in Neural Systems, 1994, 5(4): 517-548.
- [6] CHEN M J, BOVIK A C. No-reference image blur assessment using multiscale gradient [J]. Eurasip Journal on Image & Video Processing, 2011, (1): 1-11.
- [7] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a "Completely blind" image quality analyzer [J]. IEEE Signal Processing Letters, 2013, 20(3): 209-212.
- [8] MITTAL A, MOORTHY A K, BOVIK A C. No-reference image quality assessment in the spatial domain [J]. IEEE Transactions on Image Processing, 2012, 21(12): 4695-4708.
- [9] ZHANG Y, CHANDLER D M. An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes [C]// Proceedings of 2013 SPIE of Image Quality and System Performance X. San Francisco, USA: SPIE Proceedings, 2013: 266-271.
- [10] MOORTHY A K, BOVIK A C. A two-step framework for constructing blind image quality indices [J]. IEEE Signal Processing Letters, 2010, 17(5): 513-516.
- [11] SHEN J, LI Q, ERLEBACHER G. Hybrid no-reference natural image quality assessment of noisy, blurry, JPEG2000, and JPEG images [J]. IEEE Transactions on Image Processing, 2011, 20(8): 2089-2098.
- [12] LIU L, LIU B, HUANG H, et al. No-reference image quality assessment based on spatial and spectral entropies [J]. Signal Processing: Image Communication, 2014, 29(8): 856-863.
- [13] SAAD M A, BOVIK A C, CHARRIER C. Blind image quality assessment: A natural scene statistics approach in the DCT domain [J]. IEEE Transactions on Image Processing, 2012, 21(8): 3339-3352.
- [14] SHEIKH H R, WANG Z, CORMACK L, et al. LIVE image quality assessment database release 2 [C]// Proceedings of the 18th European Signal Processing Conference. Aalborg, Denmark: [s.n.], 2010: 66-69.
- [15] PONOMARENKO N, LUKIN V, ZELENSKY A, et al. TID2008 - a database for evaluation of full - reference visual quality assessment metrics [J]. Advances of Modern Radioelectronics, 2009, 10(4): 30-45.

作者简介:朱文斌(1991—),男,江苏扬州人,硕士研究生。研究方向为图像质量评价。

陈 强(1979—),男,江苏南京人,教授,博士。主要研究方向为图像处理、分析与识别。

杨曦晨(1989—),男,江苏高邮人,博士研究生。研究方向图像处理、分析与评价。

(上接第80页)

字形端正,能够给被试带来强烈的视觉冲击,吸引被试更多的关注。

参 考 文 献

- [1] 阎国利.眼动分析法在广告心理学研究中的应用[J].心理学动态,1999(4):51-53.
- [2] 张晓曼.网页文字色彩搭配的眼动研究[D].杭州:浙江师范大学,2006.
- [3] 周杨.网络广告视觉表现及优化组合[D].武汉:华中科技大学,2004.
- [4] 舒建文.谈多媒体课件制作[J].中国现代教育装备,2005(4): 14-15.
- [5] CYR D, HEAD M. Color appeal in website design within and across cultures: A multi - method evaluation [J]. International Journal of Human-Computer Studies, 2010, 68(1): 1-21.
- [6] PILLAI K G, KATSIKEAS C S, PRESI C. Print advertising: Type size effects [J]. Journal of Business Research, 2012, 65(6): 865-868.
- [7] BOGART L, TOLLEY B S. The search for information in newspaper advertising [J]. Journal of Advertising Research, 1988, 28(2): 9-19.
- [8] LOHSE G L. Consumer eye movements patterns on yellow pages advertising [J]. Journal of Advertising, 1997, 26(1): 61-73.
- [9] 蒋玉石,倪国庆,李永建,等.网页广告“靶”颜色组合识别效率的实验研究[C]// 2010年工程和商业管理国际学术会议.武汉:武汉大学,2010:117-119.
- [10] 汤晓华.网络广告中的色彩特色与应用[J].群文天地,2012(3):154-155.

作者简介:沈 艳(1990—),女,江苏泰州人,硕士研究生。研究方向为市场营销。