

一种基于内容的图像质量评价测度

王涛 高新波 张都应

(西安电子科技大学电子工程学院, 西安 710071)

摘要 鉴于传统的图像质量评价测度,如峰值信噪比,不能有效地反映人对图像的视觉感知。为此,提出了一种基于内容的图像质量评价测度;在改进基于结构相似度(structural similarity, SSIM)的图像质量测度基础上,根据图像的内容将图像分成边缘、纹理和平滑区域3部分,在每个区域又利用模糊积分融入了结构相似性的数量信息,从而充分利用了图像结构信息相似性及其在位置和数量上的融合信息来全面评价图像质量。实验结果表明,利用该测度所得到的图像质量评价结果与主观评价结果有着很好的相关性,能较准确地反映人对图像质量的主观感受。

关键词 图像质量评价 结构信息 降质图像 模糊积分

中图分类号: N911.254 文献标识码: A 文章编号: 1006-8961(2007)06-1002-06

An Objective Content-based Image Quality Assessment Metric

WANG Tao, GAO Xin-bo, ZHANG Du-ying

(School of Electronic Engineering, Xidian University, Xi'an 710071)

Abstract The traditional image quality evaluation metric, such as PSNR, cannot reflect the visual perception to the image effectively. Concerning this issue a content-based image quality assessment metric is proposed in this paper. Based on a structural-similarity-based metric (SSIM) with some modifications, our approach partitions an image into three parts: edges, textures and flat regions according the content of the image and then we apply the SSIMs to fuse each part using Sugeno fuzzy integral. The new metric combines the position and quantity information with the similarity of the image structural information and gives comprehensive evaluation to the quality of the specified image. The experiment results illustrate that the proposed metric has good correlation to the subjective perception, and can reflect the image quality effectively.

Keywords image quality assessment, structural information, distorted image, fuzzy integral

1 引言

在图像压缩、通信、存储、增强以及图像水印等应用中都涉及到图像质量的评价问题,其评价指标可以作为一种准则或测度用来校准处理系统,或用于算法优化及参数优选。图像最终是为人所观看的,故其质量最准确的评价方法是主观方法。但是主观方法在实际应用中存在诸多问题,因此人们迫切需要设计客观的评价方法来近似反映主观感受。

客观图像质量评价方法根据对原始图像的依赖

程度可以分为:全参考型、无参考型和介于两者之间的部分参考型^[1]。本文重点研究全参考型的方法,即原始图像已知且被认为是完好的参照图像。在各种全参考型客观图像质量评价方法中,均方误差(mean squared error, MSE)和峰值信噪比(peak signal-to-noise ratio, PSNR)两个指标由于计算简单,且有明确的物理意义,在实际中获得了广泛应用。但实际上这些简单的评价测度并不能很好地反映人对图像质量的主观感受。

近些年来,人们开始利用已知的一些人类视觉系统HVS(human visual system)的特性来研究图像

基金项目:国家自然科学基金项目(60202004);教育部重点项目(104173)

收稿日期:2005-08-22;改回日期:2006-03-24

第一作者简介:王涛(1979-),男,2003年获西安电子科技大学学士学位,现为西安电子科技大学电子工程学院硕士研究生。主要研究方向为数字图像质量的客观评价方法研究。E-mail: w.tao@126.com

质量评价的新方法,其中有代表性的两个模型是 daly visible differences predictor 和 sarnoff visual discrimination model^[2]。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。但这些方法与 PSNR 的效果相比并没有显著的改善^[3]。这主要是因为,基于人类视觉系统的方法是建立在许多很强的假设基础上的,存在诸多不足。Wang 在分析了上述问题的基础上,提出了一种新的观点:人眼的主要作用是为了提取视场中的结构信息,而且人类视觉系统也非常适用于这个目的。根据这一思想,Wang 等提出了基于结构相似度(structural similarity, SSIM)的图像质量测度——平均结构相似度(mean SSIM, MSSIM)^[1],实验表明此测度与主观评价有较好的相关性。不过文献[4]则认为:图像的降质程度主要是由降质图像和原图像误差的结构、位置和数量信息来决定的。为了融合更多的信息建立更为合理有效的图像质量评价测度,我们通过对图像内容的分析,利用文献[5]中的方法将其划分为边缘、纹理和平坦区域,并利用模糊积分方法融合各区域中逐像素邻域的结构相似度,给出了一种图像质量的综合评价指标。新的评价测度充分利用了图像的结构、位置和数量信息,被称为基于内容的图像质量评价方法。大量实验结果表明,该测度与主观评价的一致程度较 MSSIM 更好。

2 基于结构相似度的图像质量测度

从图像形成的角度上讲,结构信息反映了场景中物体的结构,它应独立于图像的亮度和对比度。文献[1]中将两幅灰度图像中相同位置处,对应的图像块 x, y 的亮度、对比度和结构的相似度分别定义为

$$\begin{aligned} l(x, y) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \\ c(x, y) &= \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \\ s(x, y) &= \frac{\sigma_{x,y} + C_3}{\sigma_x\sigma_y + C_3} \end{aligned} \quad (1)$$

其中, μ_x 为图像块 x 的均值,反映其亮度信息; σ_x 为 x 的方差,反映其对比度信息; $\sigma_{x,y}$ 为 x, y 的相关系数,反应其结构信息的相似性。 C_1, C_2 和 C_3 均为很小的正数,以避免分母为零或接近零时造成的奇异性。联合这 3 种相似性准则而构成的 x 和 y 的相似

性测度为

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (2)$$

其中, $\alpha > 0, \beta > 0, \gamma > 0$ 。作为相似性测度,它满足:

(1) 对称性: $SSIM(x, y) = SSIM(y, x)$ 。

(2) 有界性: $SSIM(x, y) \leq 1$ 。

(3) 唯一最大值: $SSIM(x, y) = 1$ 当且仅当 $x = y$ 取得最大值。

若取 $C_1 = (K_1L)^2, C_2 = (K_2L)^2, C_3 = C_2/2, K_1 \ll 1, K_2 \ll 1, L$ 是图像中像素灰度的动态范围, $\alpha = \beta = \gamma = 1$, 则式(2)的一种特殊的形式为

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{x,y} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

文献[1]中,首先,通过一个 11×11 的中心对称高斯加权窗 $W = \left\{ w_i \mid \sum_{i=1}^N w_i = 1, i = 1, 2, \dots, N \right\}$ (N 为窗口中的像素个数)在图像上逐点计算以该点位中心的邻域图像块的 μ_x, σ_x 和 $\sigma_{x,y}$

$$\begin{aligned} \mu_x &= \sum_{i=1}^N w_i x_i \\ \sigma_x &= \left(\sum_{i=1}^N w_i (x_i - \mu_x)^2 \right)^{1/2} \\ \sigma_{x,y} &= \sum_{i=1}^N w_i (x_i - \mu_x)(y_i - \mu_y) \end{aligned} \quad (4)$$

则由式(3)可计算出各点对应图像块的 SSIM (边缘点除外)并可映射为一幅图像来描述降质图像的降质信息,称这幅图像为 SSIM 索引图。这里采用高斯加权窗是为了避免出现分块效应。最后,用平均结构相似度作为对整体图像质量的评价测度,其中 N 为 SSIM 索引图中的点的个数。

$$MSSIM = \frac{1}{N} \sum_{j=1}^N SSIM(x_j, y_j) \quad (5)$$

3 基于内容的图像质量评价测度

为了构造更有效合理的图像质量评价标准,根据文献[4]观点:图像降质的主要因素有误差的数量、误差的位置和误差的结构。本文提出了一种基于内容的图像质量综合评价测度(content-based metric, CBM)。

CBM 测度以 SSIM 为基础,首先提取图像像素的结构信息,然后将它们按内容划分到边缘、纹理和平坦区域,再利用模糊积分获得各个区域的相似度,最后通过加权平均融合图像的结构、位置和数量信

息,对图像质量做出全面的评价结果。基于 CBM 测度的图像质量评价系统框图如图 1 所示,主要包括 3 个基本步骤:(1) 结构信息提取;(2) 区域分类;(3) 数量信息融合。

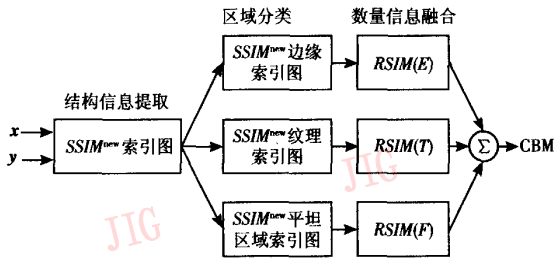


图 1 基于 CBM 测度的图像质量评价系统框图

Fig. 1 Diagram of the proposed image quality assessment system based on CBM

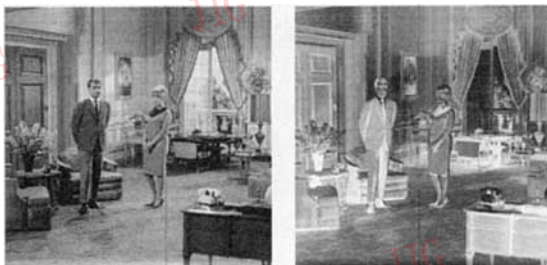
3.1 结构信息提取

这一步的主要功能是利用基于 SSIM 的方法提取被评价图像与参考图像之间的对应图像块 x, y 的结构相似度 $SSIM(x, y)$ 值并构造 $SSIM$ 索引图。

在结构相似性测度中, $l(x, y)$ 和 $c(x, y)$ 的取值范围均为 $[0, 1]$, 而 $s(x, y)$ 的值域则为 $[-1, 1]$, 不利于后续的模糊积分操作和信息融合。比如, 图 2 (b) 所示为图 2 (a) 的负片 (灰度翻转)。从图像信息表达的角度来看, 图 2 (b) 保留了图 2 (a) 的所有细节, 即结构信息。从信息传递的角度讲, 我们完全可以认为两幅图像所表达的结构信息完全相同, 它们的不同在于亮度和对比度信息。因此, 这里重新定义了图像的结构信息相似度为

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

这样, $\hat{s}(x, y)$ 的值域变为 $[0, 1]$ 。SSIM 中的另外



(a) 原图像 (b) 负片图像

图 2 结构负相关的“couple”图像结构信息比较

Fig. 2 Comparison of the structural information of “couple” images

个相似成份, $l(x, y)$ 和 $c(x, y)$ 保持不变。这样新的基于结构相似度的图像质量测度可以表示为 $SSIM^{new}(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [\hat{s}(x, y)]^\gamma$ (7) 根据式 (7) 计算计算两幅灰度图的逐点像素对 (x, y) 的 $SSIM^{new}$, 便可构成 $SSIM^{new}$ 索引图。

3.2 区域分类

这一步将一幅图像按内容划分成边缘、纹理和平坦区域 3 个部分, 分别用 R_E (下角 E 代表 edge), R_T (下角 T 代表 texture) 和 R_F (下角 F 代表 flat, 下同) 表示。在人的感知领域, 它们的重要程度依次降低, 即边缘最重要, 纹理次之, 平坦区域最低^[5]。由于 3 种区域的梯度幅度值具有依次降低的规律, 因此可以用梯度的幅度值的大小进行分类。这里仍采用文献 [5] 的方法来将图像按内容划分成上述 3 个区域:

(1) 用 Sobel 算子提取原图像和降质图像的梯度场。并分别逐点计算原图像和降质图像梯度场的幅度;

(2) 确定分割规则中所涉及的阈值为 $T_1 = 0.12gr_{max}$, $T_2 = 0.06gr_{max}$, 其中 gr_{max} 是原图像中梯度幅度的最大值;

(3) 在原图和降质图像中的对应位置上的点 (i, j) 处, 原图像的梯度幅度为 $p_o(i, j)$, 降质图像的梯度幅度为 $p_r(i, j)$, 则这点属于 R_E , R_T 和 R_F 3 个区域中的哪一个根据以下规则确定:

R_1 : 如果 $p_o(i, j) > T_1$ 或 $p_r(i, j) > T_1$, 则 $(i, j) \in R_E$;

R_2 : 如果 $p_o(i, j) < T_2$ 且 $p_r(i, j) \leq T_1$, 则 $(i, j) \in R_F$;

R_3 : 如果 $T_1 \geq p_o(i, j) \geq T_2$ 且 $p_r(i, j) \leq T_1$, 则 $(i, j) \in R_T$;

根据 $SSIM^{new}$ 索引图中各点在原图像中的对应位置所在区域即可将 $SSIM^{new}$ 索引图中的各点分别划分到边界、纹理或平坦区域, 构成 $SSIM^{new}$ 边缘索引图 E 、 $SSIM^{new}$ 纹理索引图 T 和 $SSIM^{new}$ 平坦区域索引图 F 。

3.3 数量信息融合

模糊积分理论是由 Sugeno 提出的。模糊积分是在模糊测度的基础上建立起来的, 它常常被用来融合多个信息源的模糊测度来生成综合评判^[6]。

定义 3.1 设 X 是一任意集合, B 是 X 的 Borel 场。定义在 B 上一个集合函数 g , 如果它满足以下 3 个条件, 则它是一个模糊测度:

(1) 有界性: $g(\emptyset) = 0, g(X) = 1$;

(2) 单调性: 如果 $A_1 \subset A_2$, 且 $A_1, A_2 \in B$, 则有

$g(A_1) \leq g(A_2)$;

(3) 连续性:如果 $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots \subset A_n \in$

B , 则 $g(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} g(A_n)$ 。

定义 3.2 设 (X, B, g) 为一模糊空间, $f: X \rightarrow [0, 1]$ 为一 B -可测函数, 则某一集合 $A \subseteq X$ 上的基于模糊测度 g 的 Sugeno 模糊积分定义为

$$(S) \int_A f(x) dg = \sup_{E \subseteq X} [\min(\min_{x \in E} f(x), g(A \cap E))] = \sup_{\alpha \in [0, 1]} [\min(\alpha, g(A \cap F_\alpha(f)))] \quad (8)$$

其中, $F_\alpha(f) = \{x | f(x) \geq \alpha\}$ 。

模糊测度和模糊积分的详细介绍及应用参见文献[6]。

设 $X = \{x_i | i = 1, 2, \dots, N\}$, $Y = \{y_i | i = 1, 2, \dots, N\}$, x_i 和 y_i 分别表示原图像和降质图像中相同位置对应的图像块, N 为图像区域分类中, 所划分的某一个 $SSIM^{new}$ 区域索引图 $M \in \{E, T, F\}$ 中的点数。令 $M = \{m_i = SSIM^{new}(x_i, y_i) | i = 1, 2, \dots, N\}$, 则由式(8)第 2 个等式可得, 集合 M 所对应区域的整体相似度为

$$RSIM(M) = (S) \int_M SSIM^{new}(x_i, y_i) dg = \sup_{m_i \in M} [\min(m_i, g(F_{m_i}(M)))] \quad (9)$$

其中, $F_{m_i}(M) = \{m_k | m_k \geq m_i, m_k \in M\}$, $g(K) = |K|/N$, $|K|$ 表示集合 K 中的元素个数。这样就分别得到了边界、纹理和平坦区域的相似度: $RSIM(E)$, $RSIM(T)$ 和 $RSIM(F)$ 。

可以这样理解模糊积分在这里的作用: 某个区域中的任意一点的 $SSIM^{new}$ 可以看成是评价该区域图像质量的一个信息源, 模糊积分则融合这些信息源来生成对这个区域的综合评判。而每个信息源的模糊测度表示至少具有这种评判的信息源在该区域中所占的比例。随着 $SSIM^{new}$ 增大, g 单调不增。当 $SSIM^{new}$ 与 g 的值最接近时, 取两者中较小的那个就是模糊积分最终得到评判。由此可以看出, 模糊积分实际上是对 $SSIM^{new}$ 和 g 的一种权衡。此外, 这里可以利用 $SSIM^{new}$ 和 g 的上述单调关系, 快速找到它们最接近的点, 从而减少计算量。

最后对上面得到的 3 个区域的相似度进行加权求和得到图像质量测度为

$$CBM = \sum_{M \in \{E, T, F\}} w_M \cdot SSIM^{new}(M) \quad (10)$$

其中, $w_E + w_T + w_F = 1$ 。CBM 的值域为 $[0, 1]$, 当被评价图像与原图像完全相同时取得最大值 1。

4 实验结果

为了测试 CBM 测度的合理性和有效性, 我们设计了以下 3 组测试实验: 一是所提出的客观测度与主观感知的一致性实验; 二是该客观测度的有效性实验; 三是客观测度的灵敏度测试实验。实验中计算 $SSIM^{new}$ 时采用的邻域图像块大小均为 11×11 。根据文献[1]中的给出的参数, 并将文献[5]中给出的边缘、纹理和平坦区域的重要性归一化, 本文实验中使用的参数如表 1 所示。

表 1 实验中所涉及的参数取值

Tab.1 The parameters used in experiments

K_1	K_2	L	α	β	γ	w_E	w_T	w_F
0.01	0.03	255	1	1	1	0.462	0.337	0.201

4.1 客观测度与主观感知的一致性实验

使用 LIVE 图像质量评估数据库^[7]。在这个数据库中 29 幅高分辨率的 RGB 图像被压缩成 175 幅 JPEG 图像和 169 幅 JPEG2000 图像。其中还包括它们的原图像和主观评价数据和其他详细信息。

首先, 计算各降质图像由客观测度得出的客观成绩, 并用非线性回归分析与主观评价成绩 (mean opinion score, MOS) 进行拟合^[3], 然后使用文献[3]定义的以下 3 个评价标准来衡量客观预测和主观评价的一致程度: 其一为非线性回归分析后的主客观成绩相关系数 (correlation coefficient, CC), 它反映了客观测度预测的精确性; 其二为 Spearman 等级次序相关系数 (rank order correlation coefficient, ROCC), 它反映了客观测度预测的单调性; 其三为非线性回归分析后的离出率 (outlier ratio, OR), 即经非线性回归后超出对应主观成绩 2 倍方差的客观成绩的比例, 它反映了客观测度预测的稳定性。此外, 还使用了非线性回归分析后的客观测度与主观感知的均方误差 (root mean square, RMS) 和平均绝对误差 (mean absolute error, MAE) 作为评价该算法有效性的测度。

客观测度包括 CBM 和 PSNR 用第 1 节介绍的 MSSIM 算法对该数据库中的全部图像进行测试所得到的对比实验结果如表 2 和图 3 所示。CBM 的与主观成绩的 CC、ROCC 均高于其他 3 种测度, 而 MAE、RMS 和 OR 则低于其他 3 种测度。由此可见, CBM 与主观评价结果更一致。

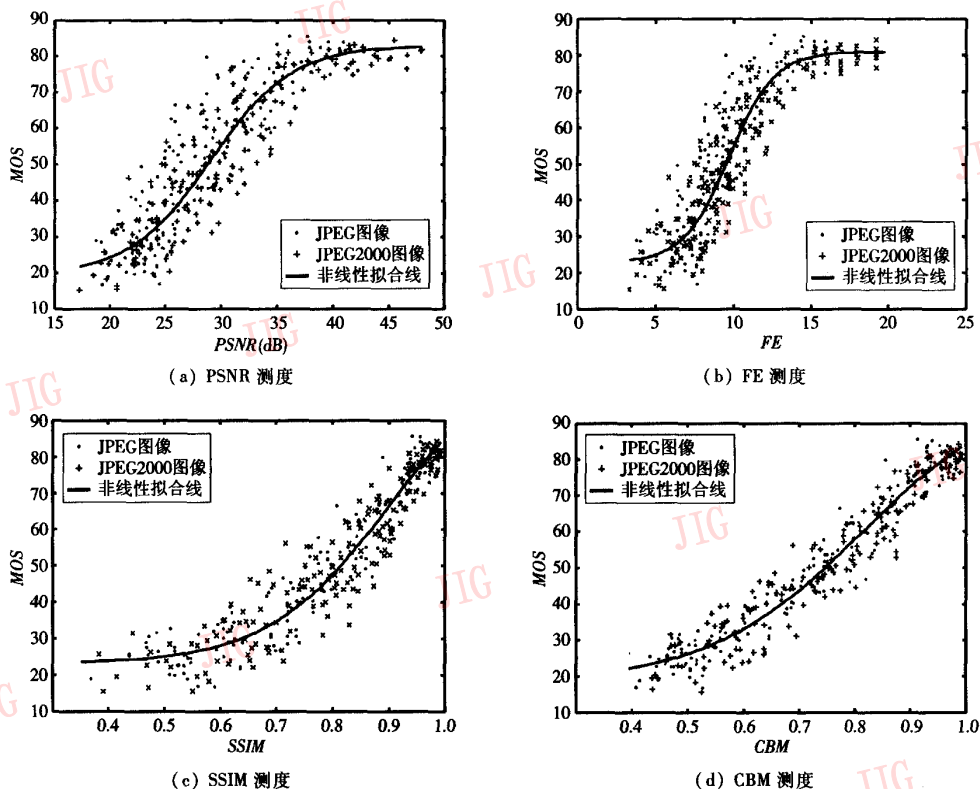


图 3 3 种图像纸质量评价测度与主观 MOS 值的关系曲线

Fig. 3 Scatter plots of subjective mean opinion score (MOS) versus model prediction

表 2 客观测度与主观感知的一致性

Tab. 2 The consistency test of CBM with MOS

Model	CC	MAE	RMS	ROCC	OR
PSNR	0.926 5	6.883	8.978	0.895 6	0.142 4
FE	0.926 9	5.445	6.799	0.893 2	0.086 0
MSSIM	0.960 6	5.432	6.745	0.944 7	0.069 8
CBM	0.975 6	4.129	5.100	0.963 1	0.011 6

4.2 客观测度的有效性实验

实验的测试图像为标准的“Lena”图像分别用窗口大小为 $W \times W$ 的平均窗模糊,加入密度为 D 的椒

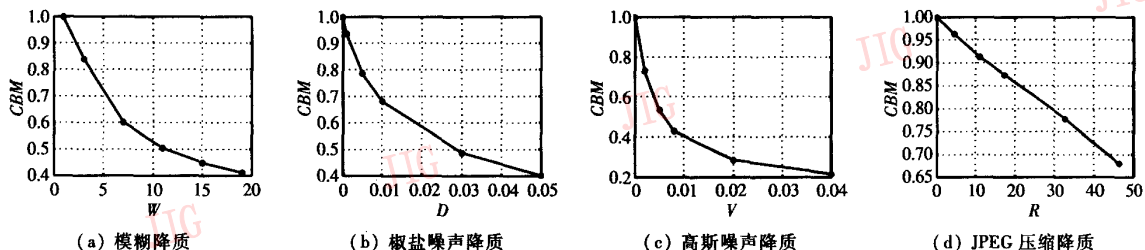


图 4 CBM 测度在不同图像降质下的变化曲线

Fig. 4 The CBM scores of “Lena” image with different types of distortions

盐噪声和加入方差为 V 的零均值高斯噪声,以及进行压缩比为 R 的 JPEG 压缩后所形成的一系列降质图像组。图 4 分别给出了随着这四种降质强度增加, CBM 逐渐下降的曲线图,这与我们所预期的一样。由此可见, CBM 对这 4 种降质类型都是有效的。

4.3 客观测度的灵敏度实验

将图 5(a) 所示的“Couple”图像经过不同处理后得到的降质图像如图 5(b) ~ 图 5(h) 所示,它们的 PSNR 值基本相同 ($PSNR \approx 26.08dB$),但从视觉上看各图像的质量明显不同。本实验以这些降质图像作为测试样本评价本文提出的 CBM 测度的灵敏

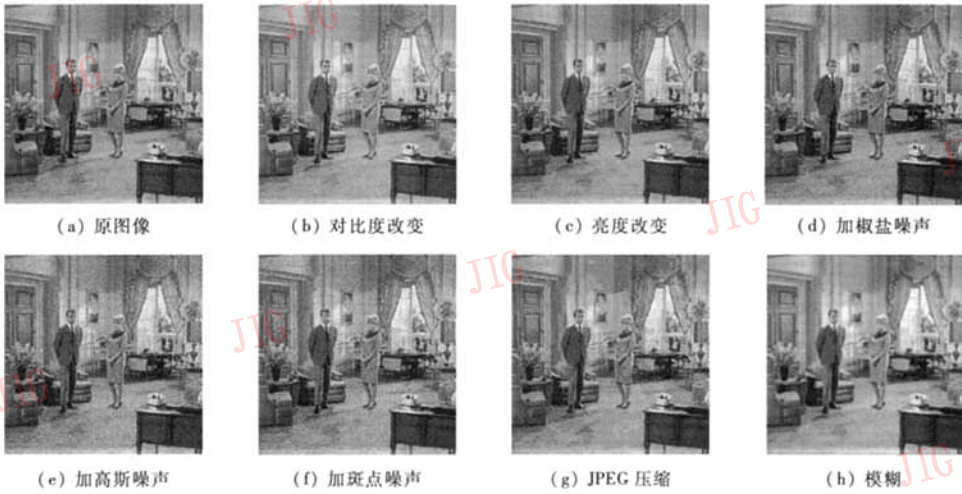


图 5 “couple” 图像在相同 PSNR 下具有不同质量的图像

Fig. 5 “couple” images with different perceived quality and the same PSNR

度。表 3 给出 CBM 测度对这些图像质量的评价值。其中由于亮度或对比度的改变后,图像的结构改变较小,因而质量明显好于经过其他降质过程的图像的质量;而 JPEG 压缩或模糊后的图像在结构上改变较大,如:出现块效应,纹理丢失等,因而其质量较差。由于 PSNR 是基于像素差异的,而 CBM 是基于结构相似性的,因而在这些情况下 PSNR 不能辨别图像质量的好坏,而 CBM 却能较好地辨别。

表 3 图 5 中所给出的各图像的 CBM 测度值

Tab. 3 The CBM scores of the images in Fig. 5

Image	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
CBM	1.00	0.961	0.961	0.749	0.713	0.710	0.667	0.643

5 结 论

本文首先分析了图像结构的负相似性相对于结构信息表达的不合理性,并对传统的 SSIM 图像质量评价测度进行了改造;然后基于图像内容分析,将其划分成边界、纹理和平坦区域,再利用降质的数量信息作为模糊测度对各区域中的结构相似度进行模糊积分,从而得到 3 个区域各自的相似度;最后对它们进行加权平均得到了最终的图像质量综合评价测度,从而将影响人感知图像质量的 3 种信息:结构信息,位置信息和数量信息有机地结合起来,以便更为精确地预测人对图像质量的主观感受。

基于内容的图像质量评价测度主要用于有原始参考图像的灰度图像质量的评价问题,由于人对彩色图像则要更为敏感,因此,如何建立针对彩色图像的质量评价测度,如何解决无参考图像的质量评价等问题,以及如何降低图像质量评价算法的复杂度问题都将是需要进一步研究的。

参考文献 (References)

- 1 Wang Z, Bovik Alan C, Sheikh Hamid R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600 ~ 612.
- 2 Li Bei, Meyer W, Klassen Victor R. Comparison of Two Image Quality Models [J]. Proceedings of SPIE Human Vision and Electronic Imaging III, 1998, 3299: 98 ~ 109.
- 3 VQEG. Final report from the video quality experts group on the validation of objective models of video quality assessment [EB/OL]. ftp://ftp.crc.ca/crc/vqeg/phase1-docs/final_report_april00.pdf, 2000, 3.
- 4 Miyahara M, Kotani K, Algazi V R. Objective picture quality scale (PQS) for image coding [J]. IEEE Transactions on Communication, 1998, 46(9): 1215 ~ 1226.
- 5 Li Jun-li, Chen Gang, Chi Zhe-ru, et al. Image coding quality assessment using fuzzy integrals with a three-component image model [J]. IEEE Transactions on Fuzzy Systems, 2004, 12(1): 99 ~ 106.
- 6 Tahani H, Keller J M. Information fusion in computer vision using the fuzzy integral [J]. IEEE Transactions on Systems, Man and Cybernetics, 1990, 20(3): 733 ~ 741.
- 7 Sheikh H R, Wang Z, Bovik A C, et al. LIVE Quality Assessment Database [DB/OL]. <http://live.ece.utexas.edu/research/quality/subjective.htm>, 2006-3-7.