

基于区域加权的视频质量评价方法

朱 宏, 郁 梅*, 唐先伟

(宁波大学 信息科学与工程学院, 浙江 宁波 315211)

摘要: 对于视频编码等视频应用来说, 准确客观地评价其视频质量至关重要. 在客观结构相似度模型的基础上, 结合人眼视觉特性提出了一种新的视频质量评价方法——基于区域加权的结构相似度视频质量评价方法. 该方法基于三维 Sobel 矩阵提取的视频时空特性, 将视频帧定义为强边缘、弱边缘和平滑三大区域进行加权求得质量分, 并结合了人眼对失真的边缘块和失真帧较为敏感等视觉特性. 在 LIVE 测试数据库中的实验结果表明, 该方法与人类视觉的主观评价结果具有较好的一致性.

关键词: 视频质量评价; 视频时空特性; 区域加权; 结构相似度

中图分类号: TN911.78

文献标志码: A

文章编号: 1001-5132 (2013) 01-0033-06

近年来, 随着互联网、通信、多媒体和视频图像处理技术的不断发展, 视频通信的应用范围也越来越广泛, 视频质量也备受关注^[1-2]. 视频质量评价主要分为主观质量评价和客观质量评价^[3]. 人眼是视觉感知的终端, 主观质量评价接近于人眼真实视觉感知, 因此较为准确可靠, 但由于费时费力、花费大、易受客观条件和主观情绪影响等缺点, 难以得到大规模的推广. 因此, 寻找简便可行的视频客观质量评价方法成为了当前的研究热点.

目前, 视频质量客观质量评价方法已有较多的模型. 较常用的是峰值信噪比(Peak Signal Noise Ratio, PSNR)和均方误差^[4-5], 此类方法原理简单、意义明确、易于计算, 但忽视了图像内容和像素间关系对人眼感知的影响, 导致客观评价结果与主观感知并不完全一致. Wang 等^[6]提出了基于结构相似度(Structural Similarity, SSIM)的视频质量评价方法(Video Structural Similarity, VSSIM), 此方法提取场景中的结构信息来模拟人眼低层次的感知. Li 等^[7]在文献[6]基础上, 基于图像内容, 将视频帧划分为边缘、纹理和平滑三大区域, 并根据人眼对不同内容的感知特性给予不同的权重, 提出基于

三分量加权结构相似度(Three-component Weighted Similarity Index, 3-SSIM)的视频质量评价方法. Wang 等人^[8]在 SSIM 基础上结合人眼感知特性, 融合对比敏感度、多通道和视觉掩蔽等信息, 提出结合感知与融合的视频质量评价方法. Liu 等人^[9]在 SSIM 基础上, 考虑到场景环境和运动信息并结合人眼对边缘信息敏感的特性, 加入边缘信息提取, 提出了基于梯度和运动估计的视频质量评价算法.

上述方法都较好地提取了空域的图像特征信息, 但没有很好地考虑到视频帧之间的时间特征, 并不能很好地评价视频质量. 因此, 笔者提出一种新型的基于区域加权的视频质量评价方法, 该方法不仅考虑到结构信息和运动信息等对人眼视觉的影响, 还考虑到人眼对边缘和失真较为敏感的特性. 实验结果表明, 笔者所提出的方法在 LIVE 视频数据库上的客观评价结果与人眼的主观评价结果一致性较高.

1 基于区域加权的视频质量评价方法

考虑到图像内容对图像质量评价的不同影响, 我们可以将 1 幅图像按其内容划分为边缘、纹理和

收稿日期: 2012-08-18.

宁波大学学报(理工版)网址: <http://nbdxlg.paperopen.com/>

基金项目: 国家自然科学基金(61271270); 浙江省自然科学基金(Y1101240); 宁波市自然科学基金(2011A610197, 2011A610200).

第一作者: 朱 宏(1988—), 男, 安徽芜湖人, 在读硕士研究生, 主要研究方向: 视频通信及图像处理. E-mail: bill_zhuhong@126.com

*通信作者: 郁 梅(1968—), 女, 江苏无锡人, 博士/教授, 主要研究方向: 视频通信及图像处理. E-mail: yumei@nbu.edu.cn

平滑区域;由于人眼对图像边缘、纹理和平滑区域具有不同的敏感度,这种规律可以由梯度幅值来表现,因此,可根据图像梯度值大小将边缘进一步分为强边缘和弱边缘,其中,弱边缘包含有失真边缘信息,容易被人眼感知^[10]。强边缘和弱边缘的梯度幅值相对较大,平滑区域的梯度幅值相对较小,而纹理区域的梯度幅值则是介于二者之间。文中为简化算法和运算复杂度,将平滑区域和纹理区域合并成平滑区域。

基于以上理论,笔者利用三维 Sobel 矩阵来提取视频帧水平、垂直和时间方向上的梯度,将视频帧分为强边缘、弱边缘和平滑区域进行加权计算视频的质量分,然后通过非线性组合来评价视频图像的优劣,图 1 给出了基于区域加权的视频客观质量评价模型的基本框架。

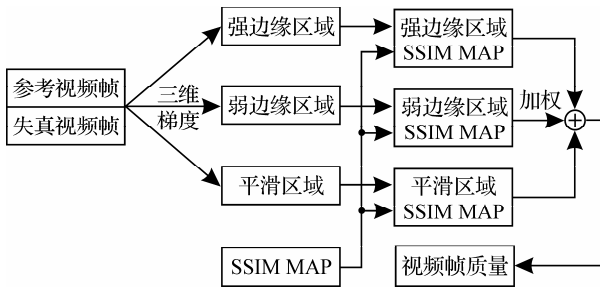


图 1 本文算法结构

1.1 基于结构相似度的视频质量评价方法

考虑到人眼主要关注视觉区域的结构信息特征, Wang 等^[6]提出基于结构相似度(SSIM)的图像质量评价方法,包括亮度 $l(x, y)$ 、对比度 $c(x, y)$ 和结构 $s(x, y)$ 等 3 个方面,

$$\text{SSIM}(x, y) = l(x, y)^\alpha c(x, y)^\beta s(x, y)^\gamma, \quad (1)$$

式中, α, β, γ 均为正数, 用于调整三部分的相关性权重。整幅图像的质量表达式为:

$$\text{MSSIM}(x, y) = \frac{1}{M} \sum_{i=1}^M \text{SSIM}(x_i, y_i), \quad (2)$$

式中, x 和 y 分别为原图像和失真图像; x_i 和 y_i 分别为原图像和失真图像中第 i 个图像块; M 为图像中块的数目。SSIM 虽然考虑了图像的结构信息, 符合人眼的部分视觉特性, 但在评价严重模糊以及混合失真图像时, 结果并不完全符合人眼主观感知, 因此不适合直接用于视频质量的评价。

1.2 区域分割

由于目前绝大多数的视频质量评价方法没有

考虑视频帧与帧之间的关系, 笔者利用三维 Sobel 模型来提取沿时间方向上的梯度。图 2 显示了沿时间方向的三维 Sobel 矩阵, 这是 1 个 $3 \times 3 \times 3$ 矩阵, 即在当前帧中, 我们需要前后 2 帧来计算其时间方向的梯度。而水平和垂直方向的梯度矩阵则是该三维矩阵分别沿 Y 轴和 X 轴旋转 90° 得到的。

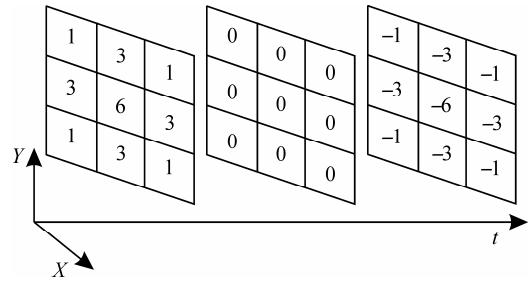


图 2 时间方向的三维 Sobel 矩阵

由于人眼对边缘区域比较敏感, 特别是针对存在失真的弱边缘块敏感的特性, 笔者将视频帧划分为 3 个区域: 强边缘、弱边缘和平滑区域。区域分割的具体算法如下:

(1) 令 g_{xr}, g_{yr}, g_{tr} 和 g_{xd}, g_{yd}, g_{td} 分别表示原始视频和失真视频沿 x, y, t 方向的梯度值, 并使用三维 Sobel 算子分别计算;

(2) 确定分割的阈值和 2 个视频的梯度幅值, 原始视频在点 (i, j) 处的梯度幅值设定为 $G_{org}(i, j)$, 并令 $G_{org}(i, j) = \sqrt{g_{xd}^2 + g_{yd}^2 + g_{td}^2}$, 同理可以确定失真视频在点 (i, j) 处的梯度幅值设定为 $G_{dis}(i, j)$, 并令 $G_{dis}(i, j) = \sqrt{g_{xd}^2 + g_{yd}^2 + g_{td}^2}$;

(3) 在原始视频和失真视频中, 对应帧中对应位置点 (i, j) 处, 原始视频帧梯度幅值为 $G_{org}(i, j)$, 失真视频帧的梯度幅值为 $G_{dis}(i, j)$, 并预定义阈值为 T , 阈值 T 定义为 $T = \beta \times G_{org, \max}$ (在平面经典算法中 β 通常为 0.12)。本文则将通过实验确定阈值 T 或参数 β , 可根据如下规则来判断: ①如果 $G_{org}(i, j) > T$, 且 $G_{dis}(i, j) > T$, 则 (i, j) 属于强边缘区域; ②如果 $G_{org}(i, j) > T$, 且 $G_{dis}(i, j) < T$; 或者 $G_{org}(i, j) < T$, 且 $G_{dis}(i, j) > T$, 则 (i, j) 弱边缘区域; ③否则即为平滑区域。

1.3 基于区域加权的结构相似度

笔者引入三维 Sobel 矩阵提取视频帧之间的时空变化特征, 所提出的算法的具体思路如下:

(1) 利用文献^[6]的方法计算整个视频序列帧按块的 SSIM map 值。

(2) 利用本算法提出的三维 Sobel 矩阵提取原始视频和失真视频的梯度, 并划分为 3 个区域: 强边缘区域、弱边缘区域和平滑区域. 分别求出 3 个区域的质量分, 那么弱边缘区域的质量分 Q_{de} 计算公式如下:

$$Q_{de} = \frac{1}{N_{de}} \sum_{(i,j) \in \text{弱边缘区域}} \text{SSIM}(i, j). \quad (3)$$

相应的强边缘区域质量分 Q_{oe} 和平滑区域质量分 Q_s 计算公式如下:

$$Q_{oe} = \frac{1}{N_{oe}} \sum_{(i,j) \in \text{强边缘区域}} \text{SSIM}(i, j), \quad (4)$$

$$Q_s = \frac{1}{N_s} \sum_{(i,j) \in \text{平滑区域}} \text{SSIM}(i, j). \quad (5)$$

(3) 给 3 个区域分别赋予不同的权值加权, 求出每帧的质量分 Q_f , 计算公式如下:

$$Q_f = a_0 \times Q_{de} + a_1 \times Q_{oe} + a_2 \times Q_s, \quad (6)$$

其中, 根据人眼视觉特性, 即人眼对边缘区域(特别是存在失真的弱边缘区域)更为敏感的特性, 确定 3 个权值分别为 a_0, a_1 和 a_2 , 并满足以下 2 个条件:

$$\begin{cases} a_0 + a_1 + a_2 = 1, \\ a_0 \geq a_1 \geq a_2. \end{cases} \quad (7)$$

(4) 对于视频中每帧的质量分采用(8)式处理, 得到整个视频的客观质量分 Q 为:

$$Q = \frac{\sum_{i=1}^F W_i \times Q_i}{\sum_{i=1}^F W_i}, \quad (8)$$

其中, F 表示视频序列中帧的数量; Q_i 是由(6)式计算的第 i 帧的得分; W_i 表示人在观看视频时很容易注意到的受损帧. 所以对质量较差的帧赋予较大的权值, 对质量较好的帧赋予较小的权值, W_i 的计算原则如下:

$$W_i = \begin{cases} 1, & 0.8 < Q_i \leq 1.0, \\ 2, & 0.6 < Q_i \leq 0.8, \\ 3, & 0.4 < Q_i \leq 0.6, \\ 4, & 0.2 < Q_i \leq 0.4, \\ 5, & 0.0 < Q_i \leq 0.2. \end{cases} \quad (9)$$

2 实验数据及实验结果

为验证本文算法, 在 Core i5 2500@3.30 GHz、

内存 8 GB 的 PC 平台上, 采用 Matlab R2010a 版本, 针对 LIVE 视频数据库^[11]进行了视频质量客观评价方法与主观感知的一致性实验. LIVE 视频质量数据库包括 10 个参考序列, 每个参考序列对应 15 个失真序列, 范围从差异不可察觉到差异极大, 共有 150 个失真视频序列. LIVE 数据库中包含有 4 种不同失真类型: MPEG-2 编码失真、H.264 编码失真、IP 传输失真和无线传输失真. 按照 VQEG 对客观质量评价方法的检验标准^[12], 利用非线性回归后客观模型预测值 $DMOS_p$ 与主观评价所得的 $DMOS$ 的相关性来度量客观评价方法的性能, 使用以下 4 个性能指标对该模型进行评价:

(1) 线性相关系数(LCC): 用 Pearson 线性相关系数 LCC 来反映客观评价模型预测的精确性, 其计算公式如下:

$$LCC = \frac{\left(\sum_{i=1}^N (DMOS_i - \overline{DMOS})(DMOS_{pi} - \overline{DMOS_p}) \right)}{\left(\sqrt{\sum_{i=1}^N (DMOS_i - \overline{DMOS})^2} \sqrt{\sum_{i=1}^N (DMOS_{pi} - \overline{DMOS_p})^2} \right)}, \quad (10)$$

其中, N 表示被测试失真立体图像的数目. Pearson 相关系数取值范围为区间 $[-1, 1]$, 其绝对值越接近 1, 表明客观模型预测值 $DMOS_p$ 与主观评价所得的 $DMOS$ 之间的相关性越好, 客观评价模型越准确.

(2) Spearman 等级相关系数($SROCC$): 用 Spearman 等级相关系数来衡量客观模型的单调性, 其计算公式如下:

$$SROCC = 1 - \frac{6 \sum_{i=0}^{N-1} (R_{DMOS_{pi}} - R_{DMOS_i})^2}{N(N^2 - 1)}, \quad (11)$$

其中, $R_{DMOS_{pi}}$ 和 R_{DMOS_i} 分别表示 $DMOS_{pi}$ 和 $DMOS_i$ 在各自样本组的排列序号. Spearman 相关系数也是区间 $[-1, 1]$ 上的 1 个值, 其绝对值越接近 1, 表明客观模型预测值 $DMOS_p$ 和主观评分差值 $DMOS$ 之间的单调性越好.

(3) 异常值比率指标(OR): 该指标主要反映客观模型的离散程度, 即预测值 $DMOS_p$ 与主观评价所得的 $DMOS$ 的差异大于某个阈值的失真立体图像数目所占的比例. OR 值得范围为区间 $[0, 1]$, 值越接近 0, 表示客观模型的一致性越好, 如果第 i 幅失真图像满足以下条件:

$$DMOS_p(i) - DMOS_i > 2S_i, \quad (12)$$

则认为第 i 幅失真图像为异常值, 其中, S_i 表示第 i 幅失真立体图像对应的 $DMOS$ 值的标准差。

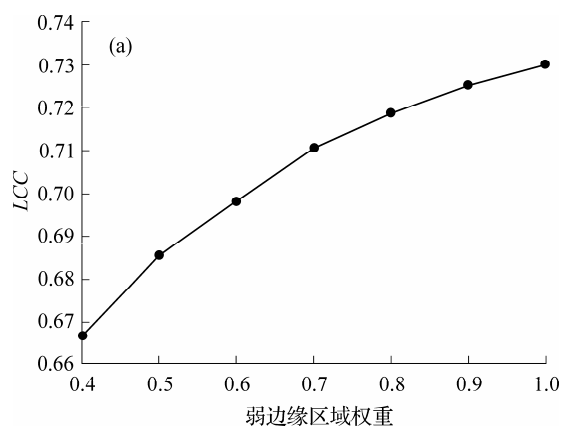
(4) 均方根误差(RMSE): 以 $RMSE$ 来对客观模型的准确性进行度量, 其值越小, 表示客观评价算法对主观评分值得预测越准确, 模型的性能越好; 反之越差。

首先来探讨基于区域加权的视频质量评价模型中的参数 T , a_0 , a_1 和 a_2 是如何确定的, 然后分析该模型与主观感知间的有效性和可靠性。

2.1 参数的确定

2.1.1 参数 a_0 , a_1 和 a_2 的确定

参数 a_0 , a_1 和 a_2 的取值大小决定各个区域的权重大小。边缘部分在视频帧感知中起到了很重要的作用, 特别是失真的边缘信息会极大地影响到感知质量。根据(7)式修改不同区域的权重, 给边缘区域特别是弱边缘区域分配更大的权重。图3显示在 a_0 , a_1 和 a_2 不同取值情况下, 150个失真视频质量评价 Q 与主观感知评价 $DMOS$ 之间的曲线



图。其中, 弱边缘区域的权重从 0.4 到 1.0 变化, 步长为 0.1, 然后将相同权重赋予强边缘区域和平滑区域, 最后根据(6)式加权得到整帧质量分。此时分区域阈值采用平面经典算法阈值, 即 $T = 0.12 \times G_{org,max}$, 其中, $G_{org,max}$ 为原始视频帧的梯度幅值的最大值。

从图3可知, a_0 , a_1 和 a_2 的取值对视频质量评价值和主观感知的相关性存在一定的影响, 特别当 a_0 取值较小时, LCC 和 $SROCC$ 的值较小。本文中, $a_0 = 1$, $a_1 = 0$ 和 $a_2 = 0$ 。

2.1.2 参数 T 的确定

由于目前大部分的分区域算法均采用平面图像分区域的经典算法中给出的阈值, 但实际应用到视频中评价效果未必最好。因此, 在平面经典算法 $T = 0.12 \times G_{org,max}$ 的基础上, 让 T 的大小从 0.06 倍到 0.45 倍变化, 步长为 0.03。图4显示了在参数 T 不同取值情况下, 失真视频质量评价 Q 与主观感知评价 $DMOS$ 之间的曲线图。此时, $a_0 = 1$, $a_1 = 0$ 和 $a_2 = 0$ 。实验中, $T = 0.18 \times G_{org,max}$ 。

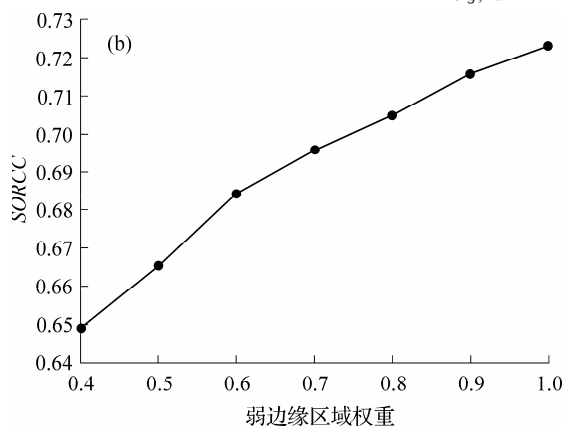


图3 在参数 a_0 , a_1 和 a_2 中, 视频质量评价与主观感知的性能曲线

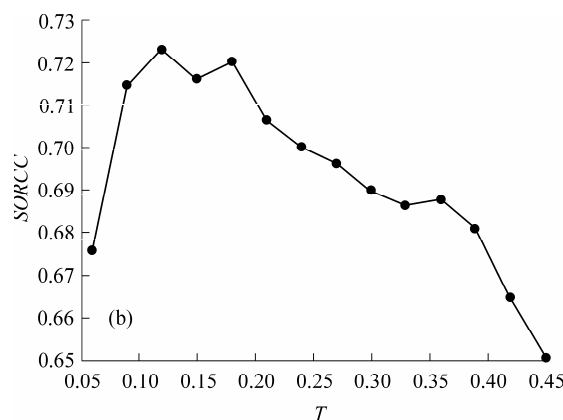
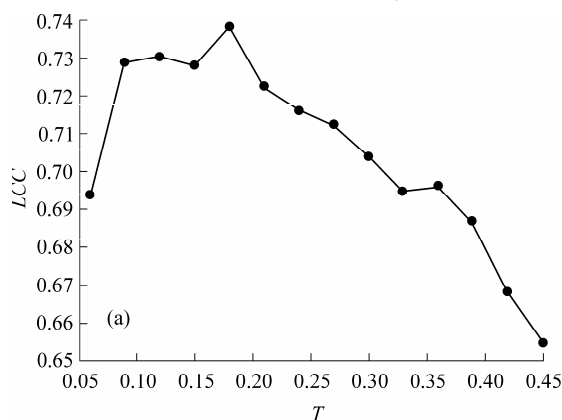


图4 在参数 T 中, 视频质量评价与主观感知的性能曲线

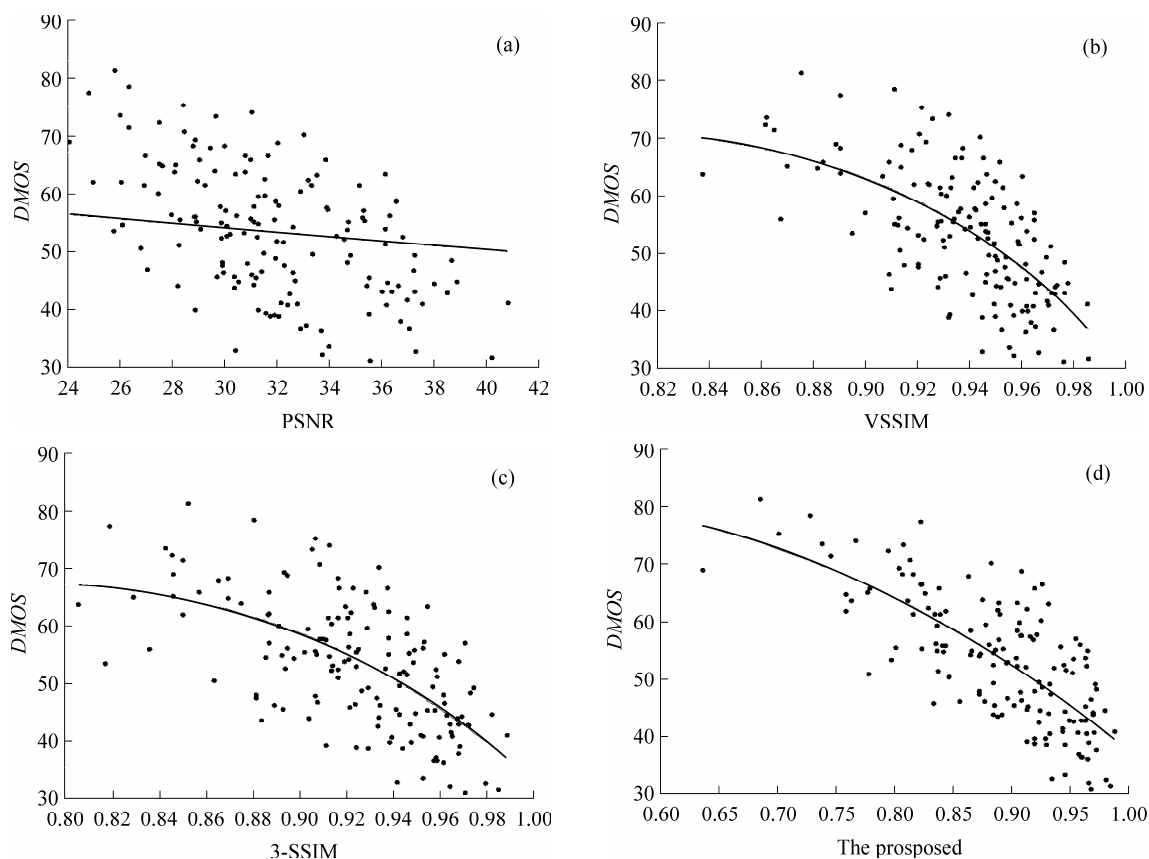


图 5 视频客观评价模型与主观感知之间的散点图

2.2 与主观感知之间的性能分析

将所提出评价算法与 PSNR^[4-5]、VSSIM^[6]和 3-SSIM^[8]进行比较, 实验结果示于图 5 和表 1. 图 5 分别给出 PSNR、VSSIM、3-SSIM 和笔者提出算法视频客观评价模型与主观感知之间的散点图. 表 1 给出了各种客观模型与主观感知的评价效果比较. 从上述数据中, 明显可以看出本文所提出的基于区域加权的视频质量评价方法与主观有很好的 consistency, 可以较好地评价视频质量.

表 1 视频客观评价模型与主观感知之间的评价性能表

算法名称	LCC	SROCC	OR	RMSE
PSNR ^[4-5]	0.539 7	0.523 3	0.1067	10.328 9
VSSIM ^[6]	0.610 5	0.600 5	0.0267	8.694 3
3-SSIM ^[8]	0.638 5	0.625 6	0.0200	8.448 7
The proposed	0.738 3	0.720 4	0	7.403 3

3 结论

由于人眼对图像的边缘信息特别是存在失真的弱边缘信息较敏感, 在 3-SSIM 算法的基础上,

将视频序列的帧图像中引入时间梯度信息, 保留了视频帧之间的时空特征信息, 并将视频帧分为强边缘、弱边缘和平滑区域, 赋予不同的加权系数以到达最佳评价效果, 使得算法更符合人眼的视觉效果. 实验表明, 提出的算法更符合人眼视觉系统特性, 能较好的评价视频质量.

参考文献:

- [1] Wang Y, Jiang T, Ma S, et al. Novel spatio-temporal structural information based video quality metric[J]. IEEE Transactions on, Circuits and Systems for Video Technology, 2012, 22(7):989-998.
- [2] Li S, Ma L, K N Ngan. Full-reference video quality assessment by decoupling detail losses and additive impairments[J]. IEEE Transactions on, Circuits and Systems for Video Technology, 2012, 22(7):1100-1112.
- [3] 蒋刚毅, 黄大江, 王旭, 等. 图像质量评价方法的研究进展[J]. 电子与信息学报, 2010, 32(1):219-226.
- [4] Rohaly A M, Coriveau P, Libert J M, et al. Video quality experts group: Current results and future directions[J]. SPIE, 2000, 4067:742-753.

- [5] Winkler S. Digital video quality: Vision models and metrics[M]. Switzerland: John Wiley and Sons, 2005:71-156.
- [6] Wang Z, Lu L, Bovik A C. Video quality assessment based on structural distortion measurement[J]. Signal Processing: Image Communication, 2004, 19(2):121-132.
- [7] Li C, Bovik A C. Content-weighted video quality assessment using a three-component image model[J]. Journal of Electronic Imaging, 2010, 19(1):11003-11008.
- [8] 王正友, 胡国胜, 吴海燕, 等. 一种结合感知与融合的视频质量评价新方法[J]. 计算机仿真, 2010, 40(10): 262-266.
- [9] 刘文娟, 李素梅, 臧艳军, 等. 基于梯度和运动估计的视频质量评价[J]. 光子学报, 2011, 27(6):1547-1552.
- [10] Ran X, Farvardin N. A perceptually motivated three-component image model, Part I: Description of the model [J]. IEEE Trans on Image Processing, 1995, 4(4):401-415.
- [11] Seshadrinathan K, Soundararajan R, Bovik A C, et al. Study of subjective and objective quality assessment of video[J]. IEEE Trans on Image Processing, 2010, 19(6): 1427-1441.
- [12] Brunnstrom K, Hands D, Speranza F, et al. VQEG validation and ITU standardization of objective perceptual video quality metrics[J]. IEEE Signal Processing Magazine, 2009, 26(3):96-101.

Video Quality Assessment Using Region-weighted Model

ZHU Hong, YU Mei*, TANG Xian-wei

(Faculty of Information Science and Technology, Ningbo University, Ningbo 315211, China)

Abstract: To evaluate the video quality accurately and objectively is important for video application, especially for video coding. Based on Structure Similarity (SSIM) model, we propose a new video quality assessment method by combining the human visual characteristic with the video quality assessment using a region-weighted model. Based on the video spatio-temporal characteristic extracted using three-dimensional Sobel matrix, the proposed method divides the video frame into strong edge region, weak edge region and smooth region, by which the frame-level scores are determined. Considering human visual characteristics such as sensitivity to the distortion block and distortion frame, the proposed method incorporates the frame-level quality scores into sequence-level quality score. The proposed method is tested on the LIVE subjective quality video database, and the results show the better performance using the proposed method in matching subjective ratings when comparing other classical methods.

Key words: video quality assessment; video spatio-temporal characteristic; region-weighted; structure similarity

(责任编辑 章践立)