

基于边缘差异的虚拟视图像质量评价方法

张艳^{①②} 安平^{*①③} 尤志翔^① 张兆杨^{①③}

^①(上海大学通信与信息工程学院 上海 200072)

^②(安徽财经大学计算机科学与技术系 蚌埠 230030)

^③(新型显示技术及应用集成教育部重点实验室 上海 200072)

摘要: 为了降低多视视频中的传输数据量,可适当减少包含纹理图及其对应深度图的多视数量,而在自由立体视频系统的终端采用基于深度图的绘制技术,生成新的虚拟视点。由于不准确的深度图估计、遮挡及合成算法等原因,合成虚拟视的边缘会出现高频噪声。对于这种特有的失真,用一般的评价方法并不能准确反映人眼的视觉感知。该文在分析失真虚拟视与原始视像素差异基础上,对各差异像素进行分类,并分配权重,提出了一种基于视觉加权的边缘差异质量评价方法,对边缘像素施加较高权重。经多视视频序列实验证明,该文方法相比于其它评价方法,能更好地预测人类视觉对虚拟视图像的主观感知。

关键词: 视频图像; 质量评价; 边缘差异; 虚拟视点; 视觉加权; 基于深度图像绘制(DIBR)

中图分类号: TN911.73

文献标识码: A

文章编号: 1009-5896(2013)08-1894-07

DOI: 10.3724/SP.J.1146.2012.01475

Quality Assessment for Virtual View Image Based on Edge Difference

Zhang Yan^{①②} An Ping^{①③} You Zhi-xiang^① Zhang Zhao-yang^{①③}

^①(School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China)

^②(Department of Computer Science and Technology, Anhui University of Finance & Economics, Bengbu 230030, China)

^③(Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai 200072, China)

Abstract: In order to reduce the transmission data in the multi-view video, the number of multi-views can be decreased properly which include texture images and depth maps, and the intermediate virtual views are generated based on Depth Image Based Rendering (DIBR) at terminal for rendering on 3D displays. Generally, the view synthesis increases the high frequency components on the edge due to inaccurate estimation of the depth map, occlusion and synthesis algorithms. Conventional 2D image quality metric is difficult to reflect the virtual view distortion. In this paper, quality evaluation method based on the edge difference is proposed. Based on the analysis of the pixel difference between virtual and original views, each difference pixel is classified and assigned a visual weight, and higher weights applied to the edge pixels. Experiments for multi-view video sequences prove that the results of this metric are in accordance with characteristic of human visual system.

Key words: Video image; Quality assessment; Edge difference; Virtual view; Visual weight; Depth Image Based Rendering (DIBR)

1 引言

自由视点电视(FTV)使用户避免了佩戴立体眼镜而可从不同角度的视点观看由多视视频合成的 3 维立体场景,满足人们对场景真实和自然再现的需求^[1,2]。多视视频可以通过多路摄像机阵列对同一场景进行多个角度同时拍摄来生成,经过多视压缩编

码后传输到用户端,用户端在解码后通过特殊的显示设备显示多个视点的立体图像。然而这种直接获取方法由于每一个摄像机形成一路视频流,因此其数据量非常巨大,必须进行高效的数据压缩之后才能进行传输。为了克服这个缺点,MPEG 提出了多视视频+深度(Multiview Video plus Depth, MVD)视频格式^[3],在 FTV 系统接收端合成虚拟视点,实现用户的任意视点选择。

相对于 2 维视频,立体视频出现了新的失真类型包括^[4]:梯形效应、鬼影、纸板效应、木偶剧场效应、阶梯效应(相邻块出现不连续的现象)等。MVD 视频格式可以减少多视视频中的数据量,其中在自

2012-11-16 收到,2013-03-25 改回

国家自然科学基金重点项目(60832003),国家自然科学基金(61172096,61171084)和上海市科委重点项目基金(10510500500)资助课题

*通信作者:安平 anping@shu.edu.cn

由立体视频系统的终端主要采用了 DIBR(Depth Image Based Rendering)技术^[5,6]。通过 DIBR 过程产生的虚拟视点出现了新的失真类型，主要表现为边缘区域的伪边缘。这是由于合成过程中存在着深度变换和视点融合等步骤，基于深度的 3D 变换的不准确及遮挡等在合成视点的边缘会出现高频噪声^[7]；并且在某些位置上，由于遮挡等原因，生成的图像会产生空洞，物体边缘出现失真。对于这种合成视点特有的失真，一些测试者需要借助原始图像成对比较来区分图像差异，然而这个是耗时间的。而一般的 2 维客观评价方法不能反映失真对图像的影响程度，从而使评价结果与人类视觉感知不符。Bosc 等人^[7,8]针对合成视点的特殊失真，从提取的图像边缘上对合成视点进行分析评价。但是这些方法的前提都是建立在用边缘提取算子对图像的边缘进行提取上的，由于边缘提取方法尚待改善，并且提取结果和边缘提取因子有很大的关系，而边缘提取因子还与纹理复杂度有关，若提取算子太大，会将人眼刚刚察觉，但不特别明显的失真伪边缘忽略掉，若提取算子太小，会产生非边缘噪声，因此评价结果带有很大的主观性。

图像由边缘、纹理及平坦区域 3 部分构成，其中边缘附近的变化最能引起人眼的视觉感知，对图像的质量起关键性作用^[9]。本文根据合成虚拟视点图像的特殊失真，重点研究虚拟视边缘位置像素的变化，对与原始图像的差异像素进行分类，提出一种基于视觉加权的边缘像素差异质量评价方法，实验结果表明本文提出的评价方法更接近于人眼的视觉感知。

2 合成虚拟视的过程及失真原因

利用深度图像，DIBR 将参考图像中的像素点映射到目标图像中，如图 1 所示。DIBR 一般可以分为两步，首先将原参考视的像素 (x, y) 利用其深度信息 $d_{x,y}$ 重投影到它们对应的 3 维空间位置 (u, v, w) ；然后根据虚拟视的位置(如相机平移、旋转参数)将这些 3 维空间点再投影到虚拟相机平面进行成像得到虚拟视像素 (x_i, y_i) ^[10]。

图 2 示意了由于深度估计不准确使深度图产生失真、导致像素变换位置发生错误而使图像中的物体发生形变及位置变动。即图 2(a)中 m 点应该借助于深度图中的 M 点变换得到虚拟点 m' ，若深度估计发生错误， M 点转移到 M_d 位置上，那么 m 点就会借助于深度图中的 M_d 点变换得到虚拟点 m'_d ，那么虚拟视图图像就会如图 2(b)所示物体形状发生变形，这种失真主要体现在物体的边缘上。

为了减少遮挡造成的虚拟视的质量下降，

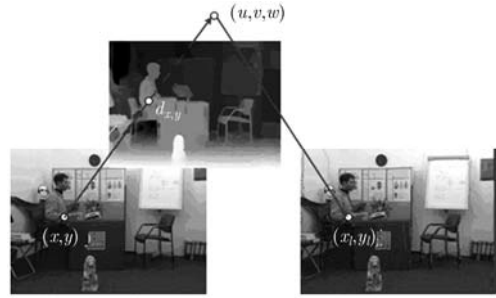


图 1 基于深度的 3D 变换

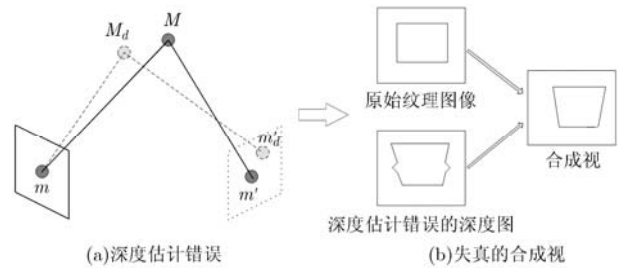


图 2 由于深度估计错误造成的合成视点失真

MPEG 小组推荐的视合成参考软件(View Synthesis Reference Software, VSRS)^[11]采用由两个邻近视图分别生成的目标图像进行融合。然而，由于深度图估计不准确、遮挡及合成算法等原因，合成视图的边缘处会造成失真，如图 3 所示。因此，评价合成虚拟视的质量需要特定的评价方法。



图 3 基于 VSRS 合成算法不同序列所得图像边缘处出现的失真

3 边缘差异评价方法

3.1 边缘差异评价方法的主要步骤

由以上分析可知，虚拟视的生成经历了基于深度的 3D 变换到合成视的生成，这样可用较少的视合成中间任意的虚拟视点。然而主要的问题是由于合成方法及深度估计的误差造成合成视中的像素误差发生是在边缘处(包括纹理区域细节部分的边缘处)，而在平坦区域很少产生误差，因此我们对边缘像素进行分析，并计算其变化情况。

其主要步骤如下：

- (1)提取原视图像与失真虚拟视图图像的差值图；

(2)人眼对像素亮度在一定区间是不易被察觉的, 设定差值图中对应的像素亮度大于阈值 T , 则失真图像对应的像素记为 $Z(i, j)=1$, 否则记为 0; 这里定义 Z 为边缘像素因子;

(3)提取 $Z(i, j)=1$ 的像素个数, 记为 n ; 那么发生失真变化的像素个数即为 n 个。

然而, 在边缘附近的边缘块是直接影响人眼的视觉感知的主要因素。以上步骤会将纹理细节区域的像素差异计算在内, 而这些区域不易引起人的关注, 因此, 我们将不易被人眼所察觉的纹理区去除掉。将上述步骤(2)进行修正, 从合成视图像中去除纹理区细节部分的差异像素。其主要步骤如下:

(a)利用 Sobel 边缘提取算子, 提取虚拟视图像的边缘;

(b)将图像分为 8×8 的图像块区域, 统计每个块区域不为零的像素个数;

(c)确定像素个数阈值 P , 若 8×8 图像块区域不为零的像素个数大于该阈值, 则认为该图像区域为细节丰富的纹理区域。计算步骤(2)所得大于阈值 T 的差值图时不考虑这些区域的像素成分。这是因为主要的差异像素为边缘处的像素, 而纹理细节区域不易被人眼感知。这里, 经实验确定阈值 P 为 16。

3.2 边缘差异评价方法的建立

合成的虚拟视点的主要失真反映在边缘附近的像素变化上, 由于边缘像素要比其它地方的像素重要, 因此我们挑选失真图像与原始图像不一样的边缘像素, 这些像素必然是失真了的像素。其它两类像素如低于阈值 T 发生变化的像素、高于阈值 T 的不易察觉的纹理细节区域像素由于不易为人眼感知, 因此这里只考虑了高于阈值 T 发生变化的边缘像素。

若原图像为 X , 失真图像为 Y , 参考 3.1 节步骤(2), 原图像和失真图像的边缘差异率为

$$\omega = 1 - \frac{\text{Num}(|Z_X \neq Z_Y|)}{M \cdot N} \quad (1)$$

其中 Z_X, Z_Y 为图像 X, Y 中边缘像素因子, $\text{Num}(|Z_X \neq Z_Y|)$ 为图像 X, Y 边缘像素因子不相同的像素个

数, M, N 分别表示图像的行数和列数。

求取与原始视图像不一样的边缘像素在一定程度上能够反映失真图像的质量, 如图 4 所示。从主观上看, 图 4(b)好于图 4(c), 圈内为图 4(c)主要发生失真的地方, 而图 4(b)未见明显失真。用本文方法和 SSIM (Structural Similarity Index Measurement)^[12]方法分别对图 4(b)和图 4(c)进行评价, 从计算得到的 ω 和 SSIM 值来看, 用 SSIM 方法不能反映主观感知, 而用本文的边缘差异率方法与主观结果一致。

本文又用同样的方法对其它合成视点进行评价, 如图 5、图 6 所示, 局部放大图为主要失真的区域, 从主观上看, 本文方法较 SSIM 方法更接近于人眼感知。

由于在设置亮度阈值 T 时, 会忽略低于阈值的像素的亮度变化, 并且阈值的设置与图像的纹理复杂度及像素的亮度有关。实验结果随阈值 T 的设置波动较大, 而限定像素个数的阈值 P 的设定不当也会在一定范围内将主要边缘处的失真像素弃除掉。因此这里我们对所有的像素加权进行计算, 减少由于阈值的设定而引起的误差。由于我们主要比较像素的亮度变化, 因此直接将边缘像素差异理论应用在像素之间的亮度差值上。这样既减少了由于本文的差异评价方法因阈值设置而造成的误差, 又考虑了人的视觉特征。

图像的评价分为 3 部分, 一部分为边缘附近发生变化的像素, 即高于阈值 T 的易被察觉的像素, 这些像素极易引起人眼的视觉感知; 另一部分是低于阈值 T 的发生变化的像素, 这些像素可能由于设置阈值的不准确而造成误差, 因此将这部分考虑在内; 还有一部分是高于阈值 T 的不易被察觉的纹理区域细节部分的像素。

若 w_i 为加权系数, Q_i 为第 i 个区域的质量, 则

$$\text{ED} = \frac{\sum_{i=1}^N w_i Q_i}{\sum_{i=1}^N w_i} \quad (2)$$

那么, 按逐像素亮度统计方法建立基于边缘差



图 4 用本文方法与 SSIM 对虚拟视点的评价结果(Champagne 序列第 40 个视点的第 300 帧)



图 5 用本文方法与 SSIM 对虚拟视点的评价结果(Champagne 序列第 38 个视点的第 300 帧)

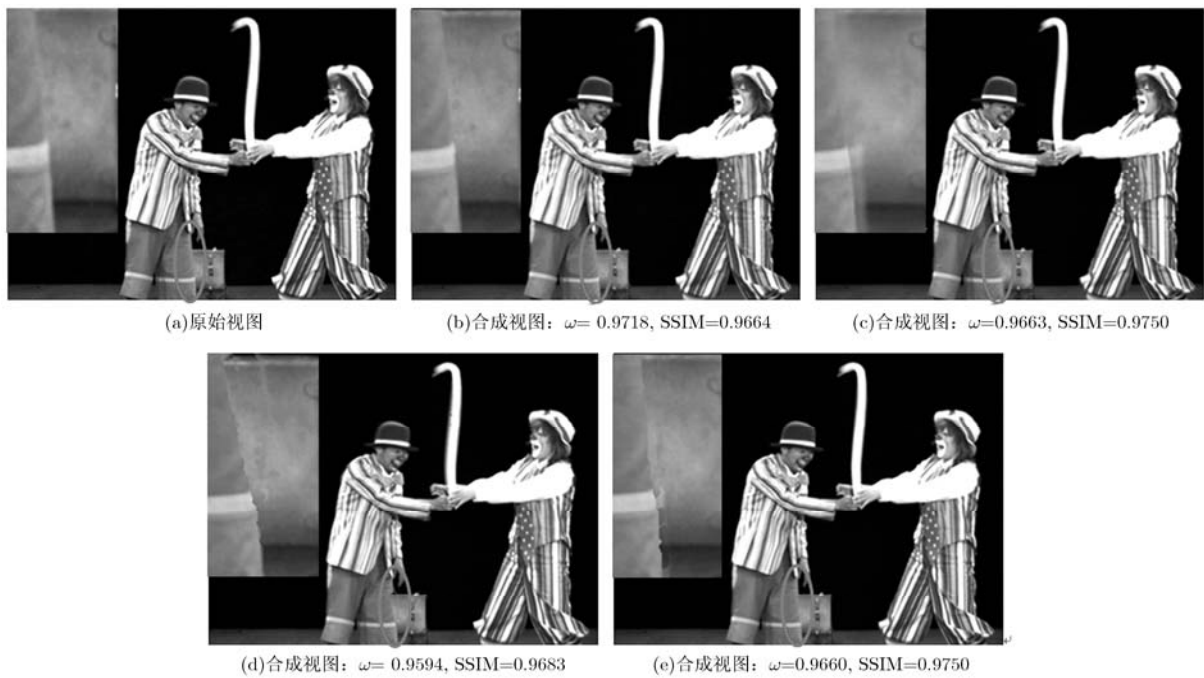


图 6 用本文方法与 SSIM 对虚拟视点的评价结果(Champagne 序列第 39 个视点的第 300 帧)

异(Edge Difference, ED)的质量评价模型为

$$\begin{aligned}
 ED &= \alpha \cdot \frac{\beta}{M \cdot N} \sum (|w_i(x_i - y_i)^\gamma|) / \sum w_i \\
 &= \alpha \cdot \frac{\beta}{M \cdot N} \left[\sum |w_{i1} \cdot (x_{i1} - y_{i1})^\gamma| \right. \\
 &\quad \left. + \sum |w_{i2} \cdot (x_{i2} - y_{i2})^\gamma| + \sum |w_{i3} \cdot (x_{i3} - y_{i3})^\gamma| \right] \\
 &\quad / \sum (w_{i1} + w_{i2} + w_{i3}) \tag{3}
 \end{aligned}$$

这里, α, β, γ 为常数, 本文分别取为: 1, 1, 2; x_i, y_i 分别为原始图像与合成视图第 i 个像素的亮度, $i \in [1, M \times N]$; x_{i1}, x_{i2}, x_{i3} 分别表示原始图像高于阈值的易察觉的像素、低于阈值发生变化的像素、高于阈值的不易察觉的纹理细节区域像素; y_{i1}, y_{i2}, y_{i3} 分别为与上述对应的合成视图的亮度; w_{i1}, w_{i2}, w_{i3} 分别为其对应的加权系数。我们经反复实验验证, 设定加权系数分别为: 0.6, 0.35, 0.05。

4 实验结果及分析

为了验证本文算法的有效性,实验采用了4组MPEG小组提供的多视视频序列Pantomime^[13], Book_arrival^[14], Lovebird1^[15], Newspaper^[16], 详细说明如表1所示。为了在合成视点图像质量分析时得到质量不同的虚拟视点图像,我们对每个序列的深度图均作模糊、扭曲及碎片失真处理,以Book_arrival和Lovebird1序列深度图为例,如图7、图8所示。虚拟视合成算法采用VSRS3.5。

表1 测试序列

测试序列	原始对(左-右)	合成的虚拟视	图像分辨率(像素)
Pantomime	37-41	38,39,40	1280×960
Book_arrival	6-10	7,8,9	1024×768
Lovebird1	4-8	5,6,7	1024×768
Newspaper	2-6	3,4,5	1024×768

为了衡量本文与主观质量之间的相关性情况,实验结果与PSNR, SSIM^[12], VQM(Video Quality Model)^[17], Bosc等人^[8]提出的方法即Bosc方法1(Canny算子阈值为0.05,标准差为1), Bosc方法2(Canny算子阈值为0.1,标准差为1)进行了比较。图9显示了5种方法与本文所提出方法的客观评价价值和主观评价价值MOS(Mean Opinion Score)的曲线拟合

图,图中所得曲线由非线性函数Logistic拟合而成,其表示式如式(4):

$$y = \frac{A_1 - A_2}{1 + (x/x_0)^p} + A_2 \quad (4)$$

其中 x, y 表示客观与主观评价价值, A_1, A_2, p, x_0 为模型参数,通过非线性函数优化求得。曲线拟合结果如图9中各图所示,横坐标的客观评价价值 Q 分别表示PSNR, SSIM, VQM, Bosc方法1, Bosc方法2及本文方法对图像的评价价值,纵坐标表示MOS值。

在建立非线性拟合之后,我们采用视频质量专家组(Video Quality Expert Group, VQEG)推荐的3个度量指标来比较各种方法的优劣^[18,19]。(1)相关系数(Correlation Coefficient, CC),反映预测的精确性;(2)Spearman等级相关系数(Spearman Rank Order Correlation Coefficient, SROCC)表征了评价方法的单调性,即预测所得值与主观值相对幅度的一致程度;(3)背离率(Outlier Ratio, OR)反映客观评价的一致性。CC和SROCC值越大,OR值越小则表明模型预测越好。本文提出的虚拟视点图像质量评价方法与其它方法的各项性能指标评价结果如表2所示,可以看出,本文方法的CC和SROCC值最大,OR值上虽然较Bosc方法2大,但是由于Bosc方法受边缘提取算子影响较大,因而本方法的准确性、单调性较好,评价更加稳定,能够更好地预测人类视觉的主观感知。



图7 Book_arrival深度图及生成的虚拟视((a),(b),(c),(d)分别表示原始、模糊失真、扭曲失真和碎片失真的深度图; (e),(f),(g),(h)分别为与(a),(b),(c),(d)对应的合成视图)

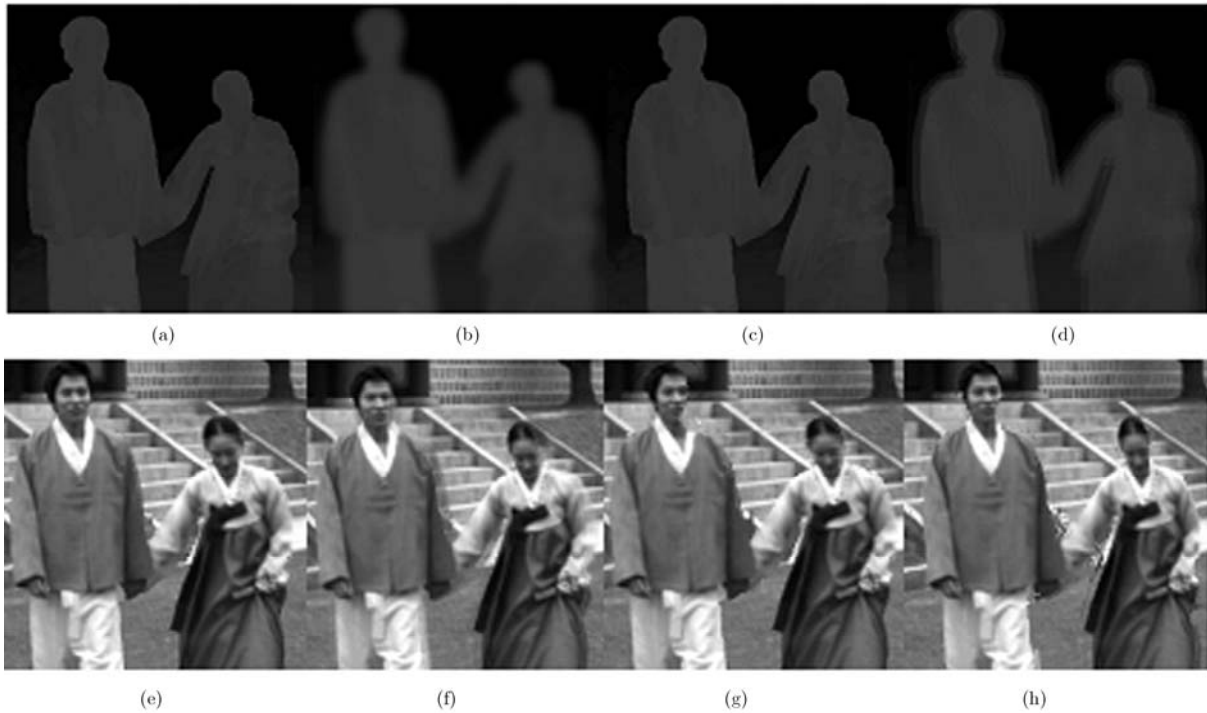


图8 Lovebird1深度图及生成的虚拟视((a),(b),(c),(d)分别表示原始、模糊失真、扭曲失真和碎片失真的深度图；(e),(f),(g),(h)分别为与(a),(b),(c),(d)对应的合成视图)

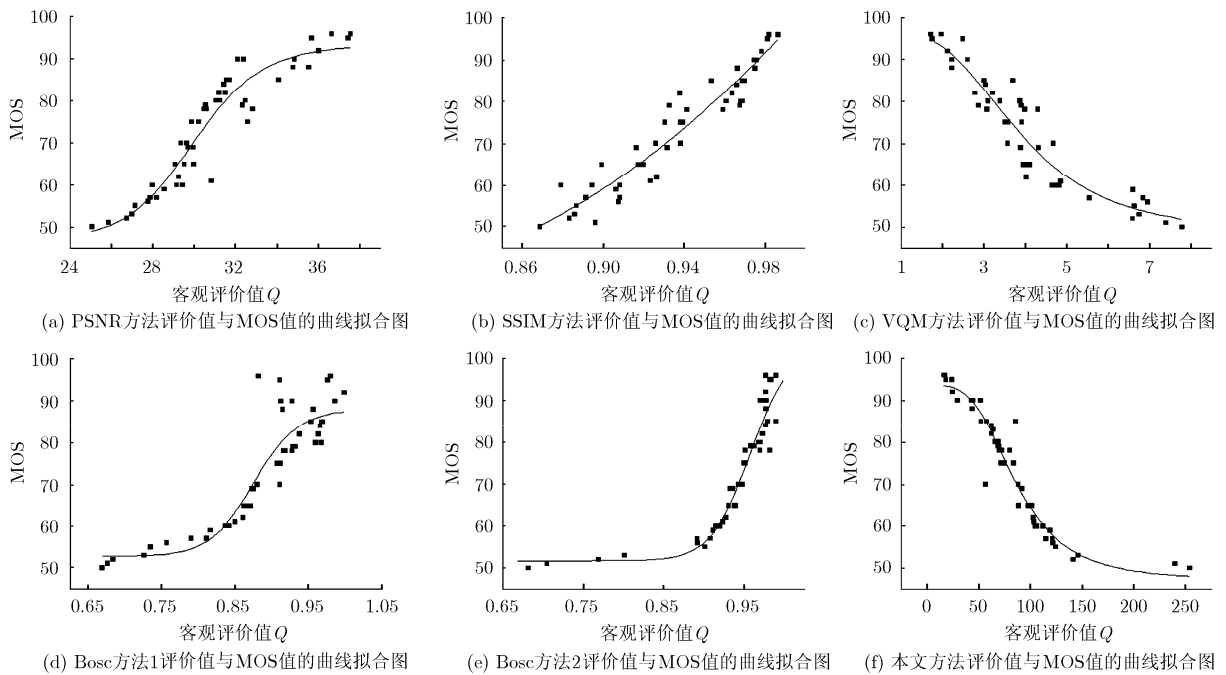


图9 各种评价方法及本文方法的客观评价价值与MOS值的曲线拟合图

5 结论

由于合成虚拟视主要的失真类型不同于一般的图像失真，因此用常用的图像质量评价方法得到的结果不能完全反应人眼的主观感知。本文针对合成视的主要失真发生在物体边缘的情况，提出了一种基于视觉加权的边缘像素差异质量评价方法。利用

与原始视的像素差异，对差异像素进行分类并加以权重，对虚拟视的边缘差异部分施以最高权重，来评价虚拟视图像质量。实验结果表明，与常用的评价方法相比，本文提出的方法与人眼的主观评价结果具有更好的关联性，更加符合人眼视觉感知，能有效评价合成虚拟视的图像质量。

表2 客观评价模型性能指标比较

评价方法	CC	SROCC	OR
PSNR	0.8965	0.9368	0.3542
SSIM	0.9142	0.9566	0.3125
VQM	0.8979	0.9458	0.3542
Bosc 方法 1	0.8134	0.8858	0.4167
Bosc 方法 2	0.9301	0.9546	0.2083
本文方法	0.9339	0.9725	0.2708

参考文献

- [1] Nicolas S H, Neil A D, Gregg E F, *et al.*. Three-dimensional displays: a review and applications analysis[J]. *IEEE Transactions on Broadcasting*, 2011, 57(2): 372-383.
- [2] Müller K, Merkle P, and Wiegand T. 3-D video representation using depth maps[J]. *Proceedings of the IEEE*, 2011, 99(4): 643-656.
- [3] Smolic A, Mueller K, Merkle P, *et al.*. Multi-view video plus depth(MVD) format for advanced 3D video systems[S]. 23rd Meeting: ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-W100, San Jose, USA, 2007.
- [4] Cancellaro M, Palma V, and Neri A. Stereo video artifacts introduced by a distributed coding approach[C]. Fifth International Workshop on Video Processing and Quality Metrics (VPQM), Arizona, USA, 2010: 1-6.
- [5] Fehn C. Depth-Image-Based Rendering (DIBR), compression and transmission for a new approach on 3D-TV[C]. Proceedings in SPIE Stereoscopic Displays and Virtual Reality Systems XI, CA, USA, 2004: 93-104.
- [6] 张秋闻, 安平, 张艳, 等. 基于虚拟视点绘制失真估计的深度图帧内编码[J]. *电子与信息学报*, 2011, 33(11): 2541-2546.
Zhang Q W, An P, Zhang Y, *et al.*. Depth map intra coding based on virtual view rendering distortion estimation[J]. *Journal of Electronics & Information Technology*, 2011, 33(11): 2541-2546.
- [7] Bosc E, Pepion R, Le Callet P, *et al.*. Towards a new quality metric for 3-D synthesized view assessment[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2011, 5(7): 1332-1343.
- [8] Bosc E, Le Callet P, Morin L, *et al.*. An edge-based structural distortion indicator for the quality assessment of 3D synthesized views[C]. Picture Coding Symposium (2012PCS), Kraków, Poland, 2012: 249-252.
- [9] Ran X and Farvardin N. A perceptually motivated three-component image model — part I: description of model[J]. *IEEE Transactions on Image Processing*, 1995, 4(4): 401-415.
- [10] Oh K J, Yea S, Vetro A, *et al.*. Virtual view synthesis method and self evaluation metrics for free viewpoint television and 3D video[J]. *International Journal of Imaging Systems and Technology*, 2010, 20(4): 378-390.
- [11] Tanimoto M, Fujii T, Suzuki K, *et al.*. Reference softwares for depth estimation and view synthesis[S]. ISO/IEC JTC1/SC29/WG11, MPEG/M15377, Archamps, France, Apr. 2008.
- [12] 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.
- [13] ISO/IEC JTC1/SC29/WG11. 1D parallel test sequences for MPEG-FTV[S]. Doc. M15378, Archamps, France, Apr. 2008.
- [14] ISO/IEC JTC1/SC29/WG11. HHI test material for 3D video[S]. Doc. M15413, Archamps, France, Apr. 2008.
- [15] ISO/IEC JTC1/SC29/WG11. 3D video test material of outdoor scene[S]. Doc. M15371, Archamps, France, Apr. 2008.
- [16] ISO/IEC JTC1/SC29/WG11. Multiview video test sequence and camera parameters[S]. Doc. M15419, Archamps, France, Apr. 2008.
- [17] Pinson M and Wolf S. A new standardized method for objectively measuring video quality [J]. *IEEE Transactions on Broadcasting*, 2004, 50(3): 312-322.
- [18] Video Quality Experts Group (VQEG). Final report from the video quality experts group on the validation of objective models of video quality assessment, Phase II[R]. <http://www.vqeg.org/>, USA, August 25, 2003.
- [19] Chikkerur S, Sundaram V, Reisslein M, *et al.*. Objective video quality assessment methods: a classification, review, and performance comparison[J]. *IEEE Transactions on Broadcasting*, 2011, 57(2): 165-182.
- 张艳: 女, 1979年生, 博士生, 研究方向为立体视频质量评价。
安平: 女, 1968年生, 教授, 博士生导师, 研究方向为图像及视频信号处理、编码及应用。
尤志翔: 男, 1983年生, 博士生, 研究方向为3维视频内容生成。
张兆杨: 男, 1938年生, 教授, 博士生导师, 研究方向为数字视频处理及多媒体通信。