

基于深度的 3D 视频质量评价*

孙延,岑峰

(同济大学 电子与信息工程学院,上海 201804)

摘要: 针对 3D 视频质量评价, 基于人眼视觉系统 HVS 提出了一种新的加权 SSIM 评测方法。利用 MVD 深度视图空间和时间上的特性, 提取主观权值映射, 基于 SSIM 利用权值映射便得到了基于深度的 SSIM 加权算法 DSSIM。实验结果表明, DSSIM 比 PSNR 更趋近于 HVS, 比 SSIM 更充分地考虑了 3D 结构信息, 得到更准确的 3D 视频质量评价。

关键词: 3D 视频质量评价; 深度视图; 权值映射; SSIM; 3D 视频合成

中图分类号: TN919.81

文献标识码: A

文章编号: 1674-7720(2013)03-0045-03

Depth perception-based 3D video quality assessment

Sun Yan, Cen Feng

(School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China)

Abstract: This paper proposes a new weighting SSIM evaluation method based on human visual system(HVS) for the 3D video quality assessment. The opinion weighting map is extracted from the features of depth view on space and time of the multi-view video plus depth(MVD). The depth-weighting SSIM(DSSIM) is proposed to assess the quality of 3D video, taking the advantage of weighting map based on SSIM. Experimental results show that, the proposed DSSIM has a better performance in 3D video synthesis compared with PSNR. This algorithm is also better than SSIM, because of the full consideration of structure information of 3D Video.

Key words: 3D video quality assessment; depth view; weighting map; SSIM; 3D video synthesis

近几年, 多媒体技术的进步使得 3D 技术作为下一代传媒系统而受到了广泛的关注。全球重要的传媒广播公司均对 3D 视频越来越重视。根据视频输入数据的不同类型, 3D 视频可分为 Multi-View Video (MVV)^[1]、Multi-View Video Plus Depth (MVD)^[2-3] 以及 Stereoscopic View。其中, MVV 包含两个或多个视图, 每个视图包含一个视频序列, 它不包含几何信息和深度信息。Stereoscopic View 是 MVV 的一种, 包括两个视图。MV 由 Single-View Video Plus Depth (VPD) 合成得到^[4]。视频的纹理信息 (Texture Information) 和深度信息 (Depth Information) 在接收端被渲染成虚拟的立体视图。深度视图是一张灰度图, 包含每个像素点的深度信息, 灰度的深浅代表物体距相机的远近。在利用深度视图得到新的虚拟视图的过程中, 视频中物体的图像被转换到它们应当的位置。所形成的虚拟立体视图就像被一个水平位置平行于真实摄像机的虚拟摄像机拍摄的一样。MVD 格

式一直是 3D 视频的主要格式。

随着 3D 研究的日益深入, 如何评估 3D 视频合成以及编码传输等各个方面的评估问题也得到广泛的关注。现在, 一些针对 3D 图像和视频质量的评估方法已经在一些参考文献中被提出。YASAKETHU S 等人^[5]使用平面的视频质量评价模型来评价 3D 视频。BENOIT A 等人^[6]提出了一种融合平面质量指标和深度失真信息的 3D 视频评价模型。YANG J 等人^[7]基于图像的绝对差值, 采用一组参数来衡量图像的质量和视频的立体感。这些算法都是基于传统的平面视频质量评价算法(如分析图像边缘特征和差异性), 都集中分析 3D 静态整体图像, 用来评价 3D 图像或者视频。但是, 它们要达到很好的效果有一个重要的前提假设, 某个特定局部的质量非常差, 以至于影响到整体的主观感知效果^[8-9]。比如, 用于视频质量评价的峰值信噪比 PSNR (Peak Signal-to-Noise Ratio) 和均方误差 MSE (Mean-Square Error) 都不能正确评价 3D 视频质量。这两种评测方法在一定条件

* 基金项目: 国家自然科学基金 (60972035)

下与HVS有较大的差异性^[10]。为了趋近于HVS,设计更好的3D视频质量评价方法成为研究的一个重要目标。

对于3D视频,在产生真实3D视觉感受和不同物体的远近层次感过程中,深度信息是一个重要因素。参考文献[11]研究了在VPD合成MVD的过程中,深度视图的压缩对合成后的MVD造成的影响。由未压缩深度视图和纹理视图合成参考MVD,由压缩深度视图和纹理视图合成测试MVD,对比参考MVD和测试MVD的PSNR,证明了深度视图对MVD质量的影响至关重要。但是文献并没有对深度视图中的信息进行提取和优化,并且评价MVD的方法也过于粗糙,没有利用深度视图中的信息,如人眼通常会更加注意近景区域以及运动区域,而深度视图与这两个区域的信息息息相关。本文的研究对象正是参考文献[11]中的参考MVD和测试MVD。依据深度视图,从中提取近景信息和运动信息,组成该视频的权值映射。在此基础上提出了新的基于深度信息的加权的SSIM算法,该方法更加符合HVS特点。

1 基于深度的SSIM加权算法

MVD由多个视角的VPD合成,针对MVD视频质量的评价,参考文献[11]中采用PSNR的算法,但其主观趋近性较差。本文首先用算法简单、高效且具备空间结构信息等优点的SSIM取代PSNR对MVD进行质量评价,然后针对SSIM并未体现人眼感兴趣区域的特性,结合深度信息提取权值映射,对SSIM算法进行加权改进。

1.1 深度视图的压缩对MVD质量影响的评估

VPD合成MVD结构示意图如图1所示。两个单视图视频分别为PVD_A和PVD_B,它们均由相应的深度视频序列和纹理视频序列组成。PVD_A和PVD_B是由固定于同一水平位置、间距固定的两个摄像头拍摄而得到的。本文使用View Synthesis Reference Software(VSRS)3.0^[12]合成MVD,将未经过压缩的深度视图和纹理视图合成参考视图MVD_r,将压缩的深度视图和纹理视图合成测试视图MVD_t,MVD_r和MVD_t使用同一来源的深度和纹理视图,在相同系统中合成而得,唯一的差别在于深度视图压缩与否。合成视图的质量是通过MVD_r和MVD_t之间的PSNR来评价的。

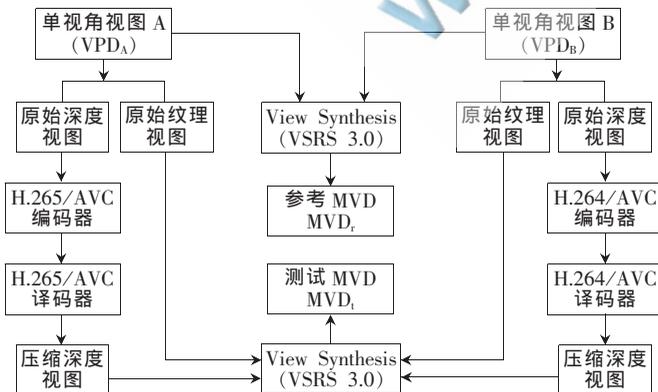


图1 VPD合成MVD结构示意图

1.2 结构相似算法(SSIM)

考虑HVS的特性,Wang Zhou等人提出的基于结构度失真的SSIM(Structural Similarity Index)评估方法被广泛应用在之后的各种模型中^[13]。SSIM方法认为自然图像信号是高度结构化的。从一个图像形成的观点来看,结构性信息是平均亮度和对比度这些反映场景中物体结构的信息。这就得到了SSIM算法的核心,即分别考量两个图像中相应块的亮度、对比度和结构信息,作为评价一个图像质量的方法。定义信号 x, y 之间的SSIM为:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_2)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

其中,信号 x, y 分别为原始块信号和失真块信号; $c_1 = (K_1L)^2, c_2 = (K_2L)^2, c_1, c_2$ 均为常数, L 为像素值的动态变化, $K_1, K_2 \ll 1; \mu_x, \mu_y$ 是亮度均值,作为亮度估计; σ_x, σ_y 是标准方差,作为对比度估计。每个块的大小通常是 8×8 。MSSIM是所有块的SSIM值加权的平均:

$$MSSIM = \frac{1}{N} \sum_{i=1}^N SSIM(x_i, y_i) \quad (2)$$

其中, N 为图像的块个数。

1.3 基于深度和运动信息的3D视频权值图

根据含深度信息的多视图3D视频的深度信息和运动信息,提出基于深度的SSIM加权测评标准,用这些信息构成权值图去衡量每个块或者每个像素点对合成视频质量的影响。由VPD得到3D视频权值图如图2所示。利用VSRS 3.0软件的矩阵变换算法将PVD_A和PVD_B的深度视图Depth_A和Depth_B变换生成Depth_{AB},Depth_{AB}包含了所有视角的深度信息,它可以代表任意视角的深度信息,即可以视为MVD_t的深度视图。

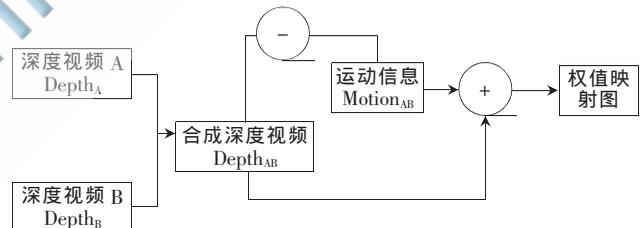


图2 由深度视图得到权值映射结构示意图

深度视图也包含物体的轮廓信息,对深度视图序列前后两帧相减即可得到物体的运动信息Motion_{AB}。为了同时考虑运动信息和深度信息,将两者加权生成该3D视频的权值图 W_{AB} :

$$W_{AB}(i, j) = \omega_M \cdot \text{Motion}_{AB}(i, j) + \omega_D \cdot \text{Depth}_{AB}(i, j) \quad (3)$$

其中, $W_{AB}(i, j)$ 是一张二维图, i, j 分别代表像素点的横轴和纵轴坐标; ω_M 和 ω_D 分别是运动信息和深度信息的加权系数, $\omega_M + \omega_D = 1$ 。根据HVS的特点,相对于空间感知,人眼对运动信息更为敏感,即 $\omega_M > \omega_D$,因此,本文将 ω_M 和 ω_D 分别设为0.65和0.35。将深度和运动这些人眼敏感的信息考虑到评价算法中, W_{AB} 可以更好地对图像模糊与失真做评估。图3为由Depth_{AB}和Motion_{AB}得到

Weighting_{AB} 的一帧示例。

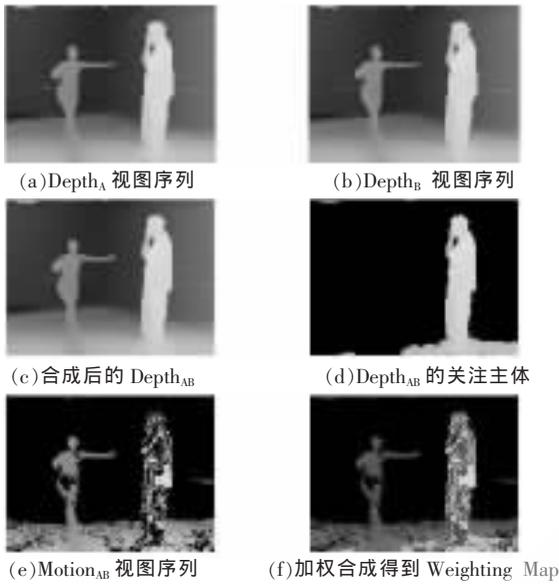


图3 由 Depth_{AB} 和 Motion_{AB} 得到 Weighting_{AB} 的一帧示例

1.4 基于深度视图的 SSIM 加权算法

3D 视频的深度信息是最重要的结构信息, 很容易依据深度信息提取出一幅图像的主体和背景。因此, 根据不同区域的权重信息, 针对深度信息的压缩对 MVD 质量影响的评估, 本文提出了更符合 HVS 特性的加权 SSIM 评价方法 DSSIM。DSSIM 基于 SSIM 算法, 考虑权值映射图像。两个输入信号分别为 MVD_r、MVD_t。由式(1)定义:

$$DSSIM(MVD_r, MVD_t, W_{AB}) = SSIM \cdot \mu_{W_{AB}} \quad (4)$$

$$\mu_{W_{AB}} = \frac{\sum_{i=l, j=h}^{i=L, j=H} W_{AB}(i, j)}{\sum_{i=1, j=1}^{i=L, j=H} W_{AB}(i, j)} \quad (5)$$

其中, $\mu_{W_{AB}}$ 是权值图中对应块的权值平均, 其中 l, h 为块的起始横、纵坐标, L, H 为块的终止横、纵坐标, I, J 分别为权值图的横、纵坐标上限。将所有块的 DSSIM 求和之后求平均, 就得到 MDSSIM:

$$MDSSIM = \frac{1}{N} \sum_{i=1}^N DSSIM(MVD_r, MVD_t, W_i) \quad (6)$$

其中, N 为图像的块个数。

2 实验结果

本文设计了一个实验, 用来验证 DSSIM 在针对含深度信息的 3D 合成视频的质量评价上, 相比于传统的 PSNR、SSIM 都有更好的效果。参考软件使用了多视角合成 VSRS 3.5, 压缩使用的是 JM15.0。

本文使用 MPEG 3DV 测试序列“ballet”(像素尺寸为 1 024×768, 帧数为 100, 帧率为 15 Hz)。图 4 是参考的 ballet 序列, 图 5 和图 6 均为压缩测试序列, 所不同的是图 5 在权值图权值高的部分(近景主体的边缘毛刺较多)产生了压缩, 图 6 在权值图权值低的部分(远景背景较模糊)产生了压缩。

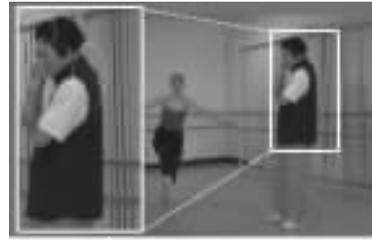


图4 无压缩 3D 合成视频序列 MVD₁



图5 测试 3D 合成视频序列 MVD₁(PSNR=34.153 493 dB; MSSIM=0.951 224 9; MDSSIM=0.937 950 5)

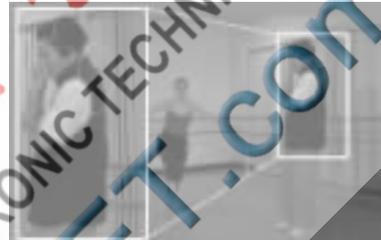


图6 测试 3D 合成视频序列 MVD₂(PSNR=33.945 160 dB; MSSIM=0.941 517 89; MDSSIM=0.941 517 8)

观察图 4、图 5 和图 6 可以明显看出, MVD₂ 的视觉效果更好, 即主观质量评价较高, 如图 5 左上的圆圈处缺少了舞者头发部分, 左下的圆圈处缺少了小臂部分, 右边圆圈处出现了毛刺等。但是依据 PSNR 和 SSIM 值, MVD₂ 的质量比 MVD₁ 要差, 这正好与主观感知相反。只有 DSSIM 值才能跟随主观视觉效果得到 MVD₂ 质量较高(DSSIM 值较大)的正确评判。

本文通过提取 3D 视频的深度和运动信息, 得到了 3D 视频的权值映射图。3D 视频中深度信息和运动信息是影响视频质量的重要因素, 从中也可以提取人眼感兴趣区域的信息。结合 HVS 特点, 继而得到了基于深度的 3D 视频客观质量评价方法 DSSIM。实验结果表明, 在传统 PSNR 和 SSIM 模型都与主观评价相悖的情况下, DSSIM 依然能够得到与主观视觉相一致的结论。3D 视频对深度感知要求较高, 基于深度的 SSIM 加权算法是一种较好的 3D 视频质量评价算法。

参考文献

- [1] TANIMOTO M. Overview of free viewpoint television [J]. Signal Processing: Image Communication, 2006, 2(6):454-461.
- [2] SMOLIC A, MULLER K, DIX K, et al. Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems [C]. Proceedings of International Conference on Image Processing, 2008: 2448-2451.

- [3] MERKLE P, SMOLIC A, MULLER K, et al. Multi-view video plus depth representation and coding[C]. Proceedings of IEEE International Conference on Image Processing, 2007 (1):1-201-I-204.
- [4] MERKLE P, WANG Y, MULLER K, et al. Video plus depth compression for mobile 3D services[C]. Proceedings of the 2009 IEEE EDTV Conference, 2009: 1-4.
- [5] YASAKETHU S L P, HEWAGE C, FERNANDO W, et al. Quality analysis for 3D video using 2D video quality models[C]. IEEE Transactions on Consumer Electronics, 2008, 54(4): 1969-1976.
- [6] BENOIT A, LECALLET P, CAMPISI P, et al. Using disparity for quality assessment of stereoscopic images[C]. ICIP 2008: the 15th IEEE International Conference on Image Processing, 2008: 389-392.
- [7] YANG J, Hou Chunping, Xu Ran, et al. New metric for stereo image quality assessment based on HVS[J]. International Journal of Imaging Systems and Technology, 2010, 20(4): 301-307.
- [8] PINSON M H, WOLF S. A new standardized method for objectively measuring video quality[J]. IEEE Transactions on Broadcasting, 2004, 50(3):312-322.
- [9] MOORTHY A K, BOVIK A C. Visual importance pooling for image quality assessment [J]. IEEE Journal of Selected Topics in Signal Processing, 2009, 3(2): 193-201.
- [10] WANG Z, BOVIK A C. Mean squared error: love it or leave it? A new look at signal fidelity measures[J]. Signal Processing Magazine, IEEE, 2009, 26(1):98-117.
- [11] EL-YAMANY N A, UGUR K, HANNUKSELA M M, et al. Evaluation of depth compression and view synthesis distortions in multiview-video-plus-depth coding systems[C]. 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video(3DTV-CON), 2010: 1-4.
- [12] TANIMOTO M, FUJIE T, SUZUKI K, et al. Reference softwares for depth estimation and view synthesis[DB/OL]. ISO/IEC JTC1/SC29/WG11, M15377, 2008.
- [13] 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.

(收稿日期:2012-10-16)

作者简介:

孙延,女,1988年生,硕士研究生,主要研究方向:视频编解码。

岑峰,男,1972年生,博士,副教授,主要研究方向:视频编解码标准的研究和整理。