

深度和结构相似性引导的四参考视点融合算法

何小梅¹, 章联军¹, 陈 芬^{1,2*}, 蔡真真², 王晓东¹

(1. 宁波大学 信息科学与工程学院, 浙江 宁波 315211; 2. 重庆理工大学 电气与电子工程学院, 重庆 400054)

摘要: 平面相机阵列四参考视点的深度图像绘制(Depth Image Based Rendering, DIBR)方案允许用户全方位身临其境地体验场景, 可有效避免虚拟视点图像边界空洞, 然而该方案引入了较为显著的伪影、背景渗透等失真。为此, 提出一种深度和结构相似性(Structural Similarity, SSIM)引导的四参考视点融合算法。首先, 深入分析了针对平面相机阵列的四参考视点 DIBR 方案中失真产生的原因; 然后, 利用参考视点与虚拟视点间的相对位置关系进行视野错误排除, 并根据恰可察觉失真模型提取融合图像的失真掩膜; 最后, 利用失真区域各视点的深度信息和 SSIM 进行自适应视点融合, 进而绘制出高质量的虚拟视点图像。实验结果表明, 本文算法绘制的虚拟视点图像比标准方案在 SSIM 和沉浸式视频峰值信噪比方面分别提升了 0.001 8 和 1.46 dB, 比文献方法在主观视觉感知方面更接近于真实图像。

关键词: 深度图像绘制; 虚拟视点融合; 平面相机阵列

中图分类号: TP391.9

文献标志码: A

文章编号: 1001-5132 (2022) 02-0096-09

沉浸式视频系统允许用户从全方位、以任意视角身临其境地体验场景, 是虚拟现实技术的发展方向之一^[1]。利用稀疏相机阵列采集的彩色图像及其深度图像^[2], 通过基于深度图像的绘制(Depth Image Based Rendering, DIBR)^[3-4]技术可实现场景重建, 是虚拟视点绘制的基本方法。针对水平相机阵列的两参考视点DIBR方案能较好地解决视点融合过程中的空洞问题, 但在绘制图像的边界处还可能存在大块空洞。近年来, 运动图像专家组(Moving Picture Experts Group, MPEG)提出了针对平面相机阵列的四参考视点DIBR方案^[5], 可有效避免绘制出的虚拟视点图像边界空洞问题, 但同时也会引入较为显著的伪影、背景渗透等失真。

Vijayanagar等^[6]根据一维邻域中非空洞像素的数量来优化左右参考视点映射图像的融合权重, 但该方法仅能改善空洞附近的失真。Lee等^[7]利用边缘信息提取出深度图的不可靠区域, 根据颜色相似性、深度可靠性和深度值进行视点融合, 可以减少伪影和模糊。Wegner等^[8]采用Z-Buffer技术对

深度差区域进行视点融合, 但该方法需要准确的深度图。Sharma等^[9]利用几何信息得到纹理匹配概率, 自适应地融合参考视点的纹理和深度信息。Qiao等^[10]采用多项式拟合方法进行视点亮度校正, 提升了虚拟视点融合准确度。蔡李美等^[11]提出一种基于分层图像的融合算法, 对前背景分割后的映射图像进行分层融合, 有效避免了背景纹理出现在前景中, 但效果取决于前背景的精确划分。上述方法都是针对水平相机阵列的两参考视点绘制方案。

在平面相机阵列四参考视点的绘制方案中, 参考视点与虚拟视点不在同一水平线上, 此时不仅存在水平视差, 而且需考虑垂直视差, 不能直接沿用两参考视点的融合方案。本文针对平面相机阵列的四参考视点DIBR方案, 提出一种深度和结构相似性(Structural Similarity, SSIM)^[12]引导的四参考视点融合算法。首先排除由视野错误导致的背景渗透失真, 并采用MPEG标准方案进行初步视点融合; 然后根据恰可察觉失真(Just Noticeable

收稿日期: 2021-07-30。

宁波大学学报(理工版) 网址: <http://journal.lg.nbu.edu.cn/>

基金项目: 浙江省自然科学基金(LY20F01000); 重庆理工大学科研启动基金(2020ZDZ029, 2020ZDZ03)。

第一作者: 何小梅(1997-), 女, 浙江金华人, 在读硕士研究生, 主要研究方向: 图像和视频信号处理。E-mail: 1367787845@qq.com

*通信作者: 陈芬(1973-), 女, 四川邻水人, 教授, 主要研究方向: 图像和视频信号处理。E-mail: chenfen@nbu.edu.cn

Distortion, JND)模型^[13]提取融合图像的失真掩膜, 进而由深度信息和图像块的SSIM自适应地确定失真区域的融合权重; 最后绘制生成高质量的虚拟视点图像. 结果表明, 本文算法绘制出的虚拟视点图像更接近于真实图像.

1 问题描述

DIBR 包括双向映射、视点融合和空洞填补三个主要环节. 图 1 示出了基于两参考视点、四参考视点绘制时各视点的位置分布. 其中, $A_i (i=1,2,3,4)$ 表示不同位置的参考视点, V 表示虚拟视点. 相比于两参考视点的绘制方案, 四参考视点的绘制方案可以避免大块空洞区域, 但生成的虚拟视点融合图像仍然会产生明显的伪影和背景渗透等失真.

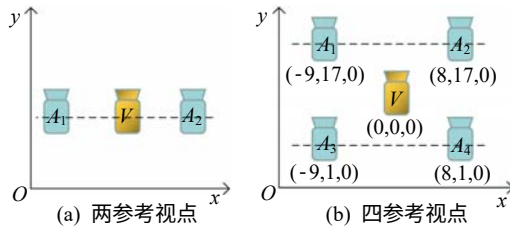
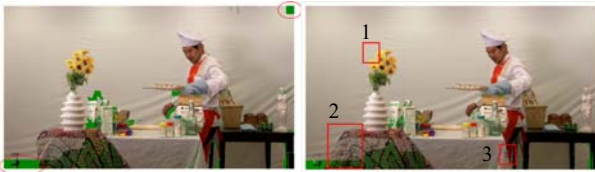


图 1 参考视点与虚拟视点位置分布

1.1 空洞分析

在基于两参考视点的DIBR方案中, 当绘制基线(图 1 中的 A_1A_2 虚线)以外的虚拟视点时, 图像边界可能存在大块空洞. 以EtriChef3 (EC3)^[2]序列为例, 采用图 1(b)所示 A_1 和 A_4 两参考视点绘制 V 视点时, 所生成的虚拟视点图像(空洞填补前)左下角和右上角存在大块空洞, 如图 2(a)所示. 在MPEG提出的基于四参考视点DIBR方案中, 采用图 1(b)中 A_1 、 A_2 、 A_3 和 A_4 四参考视点绘制 V 视点时, 可重建出绝大部分虚拟视点场景, 如图 2(b)所示. 本文统计实验结果表明, 四参考视点绘制方案重建场景的空洞面积仅为两参考视点时的 25.2%.



(a) A_1 和 A_4 绘制 (b) A_1 、 A_2 、 A_3 和 A_4 绘制

图 2 绘制基线外的虚拟视点图像(空洞填补前)

1.2 失真分析

以图 2(b)为例分析四参考视点 DIBR 方案中失

真产生的原因及现象, 图 3 示出了图 2(b)中各方框图像区域的主要失真, 图 4 为失真产生的原理示意, 其中, T_F 和 T_B 分别表示前景和背景的对象, D_F 和 D_B 分别表示前景和背景的深度.

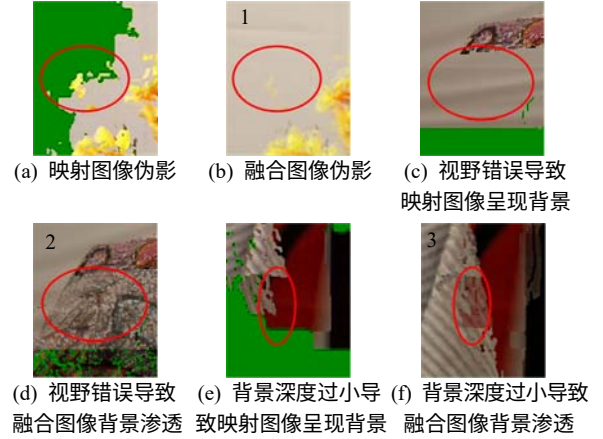
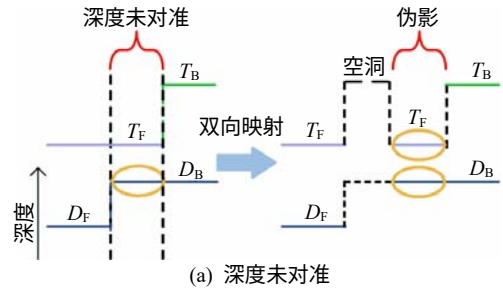
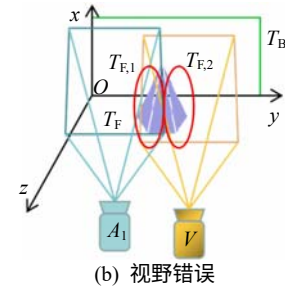


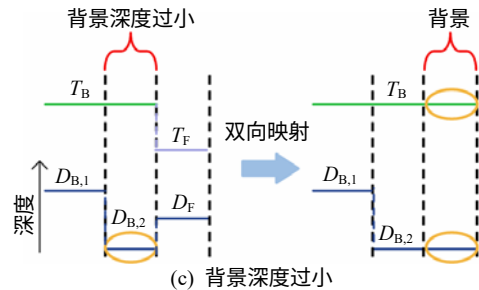
图 3 图 2(b)中的主要失真



(a) 深度未对准



(b) 视野错误



(c) 背景深度过小

图 4 失真原理示意

由于物体边缘两侧的彩色图像和深度图像之间难免存在一些未对准像素, 这可能导致伪影的产生^[14]. 图 3(a)、3(b)分别示出了图 2(b)方框 1 处

的映射图像伪影和融合图像伪影. 图 4(a)分析了此时伪影产生的原理, 当对象 T_F 、 T_B 与其深度 D_F 、 D_B 未对准时, 双向映射后, 背景对象 T_B 中出现了前景对象 T_F 的伪影.

当视点切换时, 参考视点与虚拟视点的视野不同, 本文定义为“视野错误”, 可能导致部分映射图像呈现出背景信息, 如图 3(c)所示, 且融合图像出现背景渗透失真, 如图 3(d)所示, 此时桌布的融合图像中明显存在白色背景墙渗透现象. 究其原因, 图 4(b)中 $T_{F,1}$ 和 $T_{F,2}$ 表示前景对象的两个不同区域, 由于视野错误, 通过 A_1 参考视点无法绘制出虚拟视点 V 的 $T_{F,2}$ 区域场景.

在虚拟视点绘制过程中, 深度估计不准确也可能导致映射图像呈现背景信息、融合图像出现失真. 以背景深度过小为例分析此时的背景渗透失真. 当背景深度过小时, 如图 4(c)中的 $D_{B,2}$, 它满足 $D_{B,2} < D_F < D_{B,1}$ (其中 $D_{B,1}$ 、 $D_{B,2}$ 表示两种不同的背景深度), 映射过程中 $D_{B,2}$ 对应的 T_B 覆盖 D_F 对应的 T_F , 使得部分前景对象中呈现出背景信息. 因此, 图 3(e)映射图像中明显可见部分斜条纹桌布被围裙遮挡, 图 3(f)桌布融合图像中依然可观察到围裙信息.

MPEG 提案中仅简单地将两参考视点绘制方案扩展成四参考视点绘制方案, 视点融合时采用的是 4 个参考视点分配固定权重的方法, 生成的虚拟视点融合图像仍然存在明显失真, 如图 2(b)方框 1、2 和 3 处. 由于深度估计不准确、各参考视点间存在亮度差异等原因, 可能导致不同参考视点的映射图像中亮度和色度等信息存在较大差异, 对虚拟视点融合时的贡献度不同, 因此固定权重融合法还存在一定的局限性, 本文将映射图像中这类区域称为“竞争区域”. 经统计分析可知, 四参考视点 DIBR 方案中这种竞争区域为两参考视点方案的 2.32 倍. 本文提出一种深度和 SSIM 引导的四参考视点融合算法, 克服 MPEG 提案中各参考视点固定权重融合法带来的失真, 进一步提升虚拟视点图像的绘制质量.

2 深度和结构相似性引导的四参考视点融合算法

算法流程如图 5 所示, 主要包括视野错误排除、初步视点融合、失真掩膜提取、自适应视点融

合 4 个步骤.

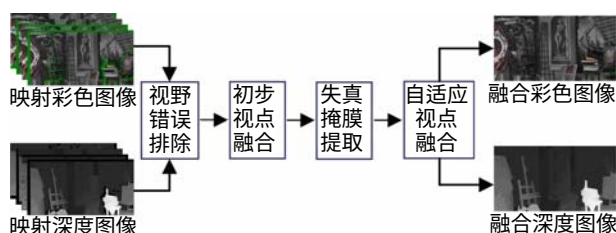


图 5 算法流程

2.1 视野错误排除

由于视野错误仅存在于参考视点与虚拟视点的图像边界处, 因此其失真位置是可预测的. 本文将图 1(b)中参考视点 A_1 与虚拟视点 V 之间的位移分解到 x 、 y 两个方向, 并在两个方向上分别进行视野错误排除, 如图 4(b)所示. 根据单应性矩阵得到 A_1 边界在 V 处对应的位置, 若 A_1 的映射图像中边界处像素的深度值大于其旁边像素深度阈值, 则判定这些像素为背景像素, 将其当作空洞处理, 从而消除因视野错误导致的背景渗透失真. 同时, 因四参考视点提供的场景内容是互相补充的, 此时并不会扩大公共空洞. 图 6(a)示出了 A_1 映射到 V 时视野错误排除后映射图像的部分细节, 对照真实场景图 6(b)可知, 图 6(a)排除了视野错误产生的背景, 即图 6(c)中暴露出的白色墙背景.

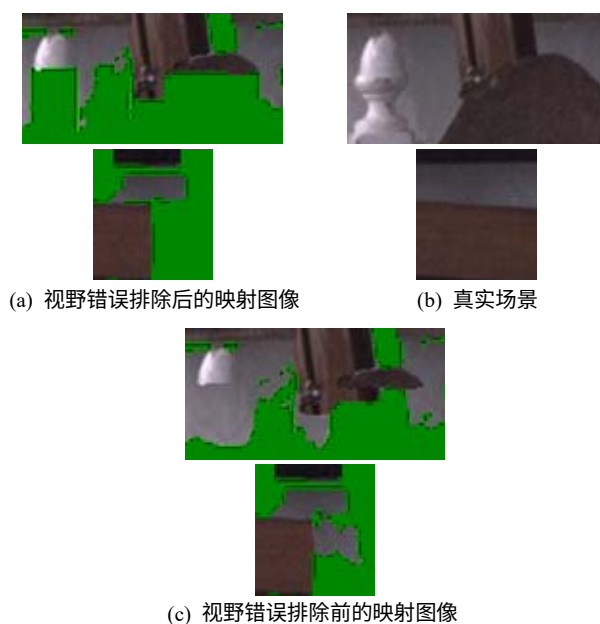


图 6 视野错误排除

2.2 初步视点融合

如前所述, 采用固定权重融合法绘制出的虚

拟图像可能存在伪影和背景渗透失真等. 因此, 在视野错误排除的基础上, 首先采用固定权重融合法进行初步视点融合, 然后提取出失真掩膜, 最后结合深度信息和图像块的SSIM^[12]自适应地对失真区域进行再次融合, 以消除相应失真.

参照MPEG提案, 对各参考视点的映射图像采用固定权重融合法(即距离的倒数)进行加权融合^[5], 表达式为

$$I_V(p) = \sum_{i=1}^4 w_{A_i}(p) \times I_{A_i}(p), \quad (1)$$

其中: $I_V(p)$ 为融合图像在 p 处的像素值; $I_{A_i}(p)$ 为对 p 有贡献的参考视点 A_i 相应的映射图像像素值; $w_{A_i}(p)$ 为 A_i 与虚拟视点 V 之间距离的倒数在 p 处的归一化权重, 计算如下:

$$w_{A_i}(p) = \begin{cases} \frac{1}{d_{A_i}} / \sum_{j \in J} \frac{1}{d_{A_j}}, & I_{A_i}(p) \neq \text{hole}, \\ 0, & I_{A_i}(p) = \text{hole}, \end{cases} \quad (2)$$

其中: d_{A_i} 为 A_i 与 V 的欧氏距离; J 为映射图像 p 处非空洞的参考视点集合.

2.3 失真掩膜提取

JND模型可以计算出人眼的不可察觉范围, 人眼对亮度域的变化比对色度域更敏感, 因此本文引入基于背景亮度掩蔽的JND模型进行失真掩膜的提取, 具体描述为^[13]

$$J_{\text{nd}}(p) = \begin{cases} 17 \times (1 - \sqrt{l(p)/127}) + 3, & l(p) < 127, \\ 0.3 \times (l(p) - 127) / 128 + 3, & l(p) \geq 127, \end{cases} \quad (3)$$

其中: $J_{\text{nd}}(p)$ 为人眼对亮度变化的最大容忍度; $l(p)$ 为 p 处的平均背景亮度, 由 5×5 大小的高斯核计算.

在四参考视点DIBR方案中, 各参考视点的映射图像中存在较多竞争区域, 因此, 本算法先提取出各参考视点的竞争掩膜. 若参考视点 A_i 的竞争掩膜记为 M_{A_i} , 将其与融合图像的绝对差值超过人眼不可察觉范围的像素标记为1, 其他像素标记为0, 则:

$$M_{A_i}(p) = \text{sgn} \left(\left| I_V(p, 1) - I_{A_i}(p, 1) \right| - J_{\text{nd}}(p) \right), \quad (4)$$

其中: $\text{sgn}(\cdot)$ 为阶跃函数; $I_V(p, 1)$ 和 $I_{A_i}(p, 1)$ 分别为融合图像和参考视点 A_i 映射图像的 p 处亮度通道像素值.

融合图像中总的失真掩膜 $M_V(p)$ 由4个参考视

点的 M_{A_i} 合并得到, 失真区域的像素标记为1, 其他像素标记为0, 则:

$$M_V(p) = \text{sgn} \left(\left| \sum_{i=1}^4 w_{A_i}(p) \times I_V(p, 1) - I_{A_i}(p, 1) \right| \times M_{A_i}(p) - J_{\text{nd}}(p) \right). \quad (5)$$

用户体验虚拟现实场景时, 对虚拟视点图像中物体边缘的轻微位移失真不敏感, 容易察觉平坦区域内的大块失真. 结构张量可以有效快速区分平坦区域、边缘区域和角点^[15], 因此本文用结构张量来计算初步视点融合图像(图7(a))的平坦区域. 每个像素的结构张量为

$$\mathbf{Z} = \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix}, \quad (6)$$

其中, g_x 和 g_y 为融合图像在水平和竖直方向上的梯度. 结构张量的迹小于50的区域近似被认为平坦区域^[15], 即图7(b)的白色区域. 本文仅提取位于平坦区域的失真掩膜, 将 $M_V(p)$ 中位于非平坦区域的像素标记为0. 用形态学膨胀操作恢复失真掩膜的边缘, 得到最终的融合图像失真掩膜, 如图7(c)所示, 其中, 白色代表失真区域.



(a) 初步视点融合图像



(b) 平坦区域分解



(c) 失真掩膜

图7 失真掩膜的提取

2.4 自适应视点融合

由于人类视觉系统(Human Visual System, HVS)只能感知由潜在的生理和心理机制决定的某个可见阈值以上的颜色变化^[13], 融合图像中不同区域可采用不同的视点融合策略, 即使生成的亮度、色度信息存在差异也不能被人眼察觉到, HVS擅长于从视觉场景中提取结构信息和颜色信息. SSIM可结合亮度、对比度和结构3个参数, 提供感知图像质量的良好近似^[12]. 因此, 本文对融合图像的非失真区域直接选取初步的视点融合结果, 而对失真区域则根据深度信息和图像块的SSIM进行自适应视点融合.

首先, 基于 9×9 大小的图像块计算各参考视点与融合图像块在 p 处的SSIM^[12]. 其中, 融合图像块是以 p 为中心在融合图像上取得的, 参考图像块是以 p_{A_i} 为中心在 A_i 参考图像上取得的, p_{A_i} 是 p 利用逆向3D-Warping在 A_i 参考图像上的对应像素. 在融合图像块和参考图像块上公共非空洞区域 Ω_{com} 内计算各像素 p_K 的结构相似性 $S_{A_i}(p_K)$ ^[12]. $\bar{S}_{A_i}(p)$ 为 p 处 A_i 参考图像块与融合图像块的SSIM^[12], 计算如下:

$$\bar{S}_{A_i}(p) = \frac{1}{|\Omega_{\text{com}}|} \sum_{p_K \in \Omega_{\text{com}}} S_{A_i}(p_K). \quad (7)$$

然后, 对失真掩膜 M_V 中的所有像素按以下步骤进行融合.

Step 1 计算 p 处各参考图像块的 $\bar{S}_{A_i}(p)$ 和所有参考视点的最大深度差;

Step 2 判断最大深度差是否超过阈值 α_d , 是则跳转 Step 3, 否则跳转 Step 4;

Step 3 判断深度值最小的参考图像块与融合图像块间的SSIM是否超过阈值 α_f , 是则选择深度值最小的视点像素值, 然后跳转 Step 7, 否则跳转 Step 4;

Step 4 统计 $\bar{S}_{A_i}(p) \geq \alpha_s$ 的参考视点数量 N , 观察是否满足 $N \geq 1$, 是则跳转 Step 5, 否则跳转 Step 6;

Step 5 将满足 $\bar{S}_{A_i}(p) \geq \alpha_s$ 的视点的图像块SSIM归一化作为权重进行融合, 跳转 Step 7;

Step 6 将所有视点的图像块SSIM归一化后作为权重进行融合;

Step 7 更新失真掩膜.

图8(a)~(d)分别示出了图1(b)中 A_1 、 A_2 、 A_3 和 A_4 各参考视点融合图像中每个像素融合时所贡献的权重, 图8中的4幅子图是互补的, 并且在除公共空洞外的任何给定像素处权重总和为100%.

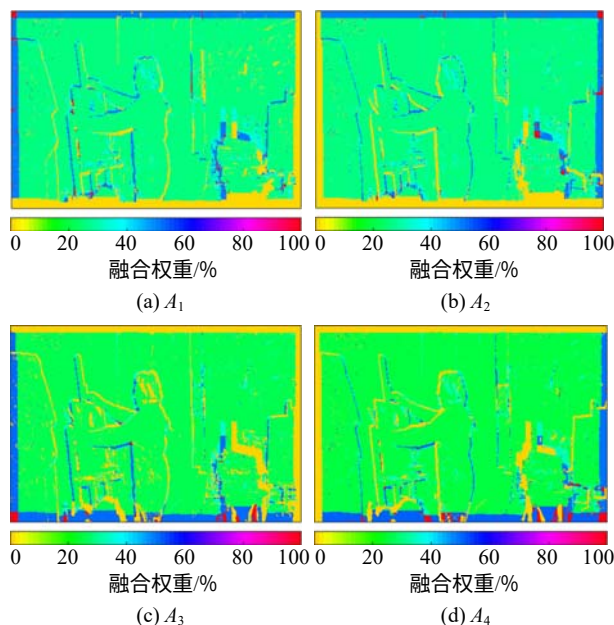


图8 各参考视点的融合权重

3 实验结果与分析

为了评估本文虚拟视点融合算法的有效性, 选取MPEG标准绘制方案的视点合成参考软件(View Synthesis Reference Software, VSRS) 4.3版^[5]和Vijayanagar算法^[6]进行对比分析. Vijayanagar算法由于其在速度和质量上的改进, 是MPEG双视点标准绘制方案的模式之一. 本实验采用TechnicolorPainter (TP)^[2]和EC3^[2]两个序列进行验证, 其彩色图像均由真实场景拍摄得到, 深度图像来源于深度估计软件DERS^[16]. 为定量比较不同算法的客观性能, 每个序列选取两个存在真实场景的视点位置进行绘制, 序列的基本参数和各视点的选取见表1, $x_m y_n$ 表示第 m 列第 n 行的视点.

3.1 客观质量对比分析

借助普通图像质量度量指标SSIM^[12]将本文算法与VSRS^[5]、Vijayanagar算法^[6]进行定量比较, 表2统计了各自视点融合图像和虚拟视点绘制图像的平均SSIM. 可见, 本文方法在SSIM上均有一定提升. 不过有部分虚拟视点图像的SSIM值不够理想, 逐帧检查绘制视点时发现, 其绘制图像几乎与真

表1 测试序列

序列	帧数	分辨率	相机阵列	相机间距	虚拟视点	参考视点
TP	372	2 048×1 088	4×4	7 cm×7 cm	x_2v_1	$x_1v_0, x_3v_0, x_1v_2, x_3v_2$
					x_1v_2	$x_0v_1, x_2v_1, x_0v_3, x_2v_3$
EC3	300	1 920×1 080	5×5	10 cm×10 cm	x_2v_2	$x_1v_1, x_3v_1, x_1v_3, x_3v_3$
					x_2v_3	$x_1v_2, x_3v_2, x_1v_4, x_3v_4$

表2 融合图像和虚拟视点图像平均 SSIM 对比

测试序列	融合图像			虚拟视点图像		
	文献[5]	文献[6]	本文	文献[5]	文献[6]	本文
TP_ x_2v_1	0.935 2	0.931 4	0.936 7	0.938 2	0.934 5	0.938 9
TP_ x_1v_2	0.927 8	0.923 5	0.929 3	0.931 8	0.927 6	0.932 1
EC3_ x_2v_2	0.912 6	0.904 3	0.919 9	0.926 0	0.917 7	0.927 6
EC3_ x_2v_3	0.874 3	0.865 9	0.882 9	0.885 3	0.876 9	0.889 8
平均增益	-	-0.006 2	0.004 7	-	-0.006 2	0.001 8

实图像完全相同,这是由于前景像素在映射时产生了 2~3 个像素的位置偏移,偏移会影响SSIM值的计算,但不会影响用户的视觉观看体验.对此,MPEG提议采用沉浸式视频峰值信噪比(Immersive Video Peak Signal to Noise Ratio, IVPSNR)^[17]来度量虚拟视点图像的客观质量.本文将IVPSNR性能指标命名为 Q ,它在YUV颜色空间中的计算式为

$$Q = \sum_{c=0}^2 10 \lg \left(\frac{R_{\text{MAX}}^2}{E(c)} \right) \cdot w_{\text{ccw}}(c) / \sum_{c=0}^2 w_{\text{ccw}}(c), \quad (8)$$

式中: $c \in \{0, 1, 2\}$, 依次对应 Y、U、V 三个通道; 各通道的颜色分量权重 $w_{\text{ccw}}(c)$ 分别为 1、0.25、0.25; R_{MAX} 为颜色分量的最大值; $E(c)$ 为考虑了像素位移和全局色差的 c 通道的均方误差, 计算如下:

$$E(c) = \frac{1}{W \cdot H} \sum_p \min \left\{ I_s(p, c) - I_G(p_G, c) + \max \left\{ \frac{1}{W \cdot H} \sum_p [I_G(p, c) - I_s(p, c), R_{\text{MUD}}(c)] \right\}^2 \right\}, \quad (9)$$

式中: $R_{\text{MUD}}(c)$ 为 c 通道颜色分量的最大不可见差异, 取值为 1%; W 和 H 为图像的宽度和高度; p_G 为以 p 为中心的 5×5 窗口内的像素; $I_s(p, c)$ 和 $I_G(p, c)$ 分别为虚拟视点绘制图像和真实图像中 p 处 c 通道的像素值.

表 3 统计了本文方法所得到的视点融合图像

和虚拟视点绘制图像的IVPSNR, 本文算法优势明显, 且其平均IVPSNR值大于 38 dB, 这表明本算法所绘制的虚拟视点图像和真实图像之间具有高度相关性. 可见, 虚拟视点绘制图像的质量评价指标中, IVPSNR与SSIM相比, 可产生更可靠的评价结果. 并且, 相比于VSRS^[5]生成的虚拟视点图像, Vijayanagar算法^[6]在SSIM和IVPSNR上质量都降低了, 而本文算法在SSIM和IVPSNR上表现更优, 分别取得了 0.001 8 和 1.46 dB的性能增益.

表3 融合图像和虚拟视点图像平均 IVPSNR 对比 dB

测试序列	融合图像			虚拟视点图像		
	文献[5]	文献[6]	本文	文献[5]	文献[6]	本文
TP_ x_2v_1	35.19	35.14	36.63	46.07	45.70	46.82
TP_ x_1v_2	34.18	34.14	35.53	45.86	45.82	46.16
EC3_ x_2v_2	26.12	26.07	28.82	38.35	37.86	39.92
EC3_ x_2v_3	26.70	26.30	28.83	35.58	34.16	38.82
平均增益	-	-0.14	1.91	-	-0.58	1.46

通过对比分析所绘制的虚拟视点图像在帧级上的客观质量, 统计结果表明, 本文算法在 96%以上的帧级质量都明显优于VSRS^[5]和Vijayanagar算法^[6], 如图 9 所示. 结果还表明这些算法所绘制的 TP_ x_2v_1 、EC3_ x_2v_3 虚拟视点图像的帧级SSIM和IVPSNR都存在一定波动, 因为在深度估计过程中每一帧都是独立估计的, 导致各参考视点的每一帧深度图质量起伏不定, 而场景又是动态变化的, 因此造成所绘制虚拟视点图像的质量不平稳.

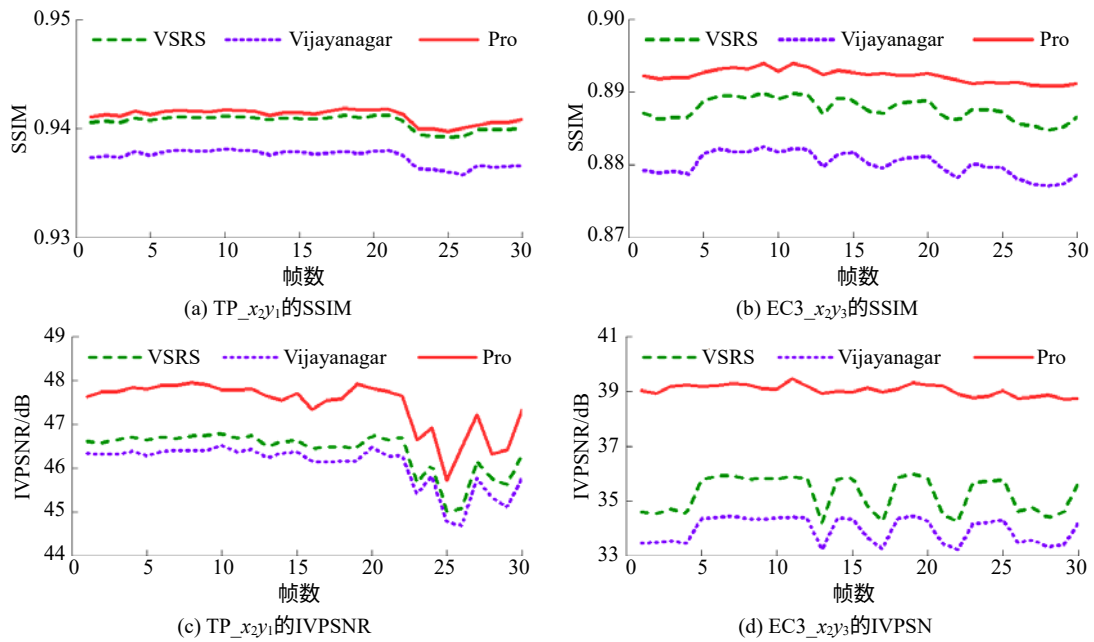


图9 帧级客观质量对比

3.2 主观质量对比分析

图10左侧展示了本文方法重建 $TP_{x_2y_1}$ 第118帧和 $EC3_{x_2y_3}$ 第60帧图像的整体结果,左上角标注了其IVPSNR,右侧是细节放大图像,分别对应VSRS^[5]、Vijayanagar算法^[6]、本文算法结果和真实图像。可见,本文算法较好地消除了伪影、背景渗透失真等,更接近于真实图像,尤其是近景物体、图像纹理边缘等区域,重建出的场景更加合理,主观视觉效果更舒适;Vijayanagar算法^[6]在前景与背景发生竞争时倾向于选择大块的背景纹理;

VSRS^[5]难以避免前景和背景相互扩散,导致产生了一些与真实场景不符的模糊区域。

本文算法对不同的虚拟视点提升效果不一致,这是因为所绘制的虚拟视点图像质量的提升幅度取决于该视点失真区域的像素数量。从图10可见,本文算法显著提高了 $EC3_{x_2y_3}$ 图像的质量,尤其是图中纹理丰富的桌布和桌布左侧区域,但对 $TP_{x_2y_1}$ 图像的质量提升相对有限,这是由于当深度图分布较集中时,各参考视点竞争导致的失真区域面积较小。图11示出了两个视点的深度直方

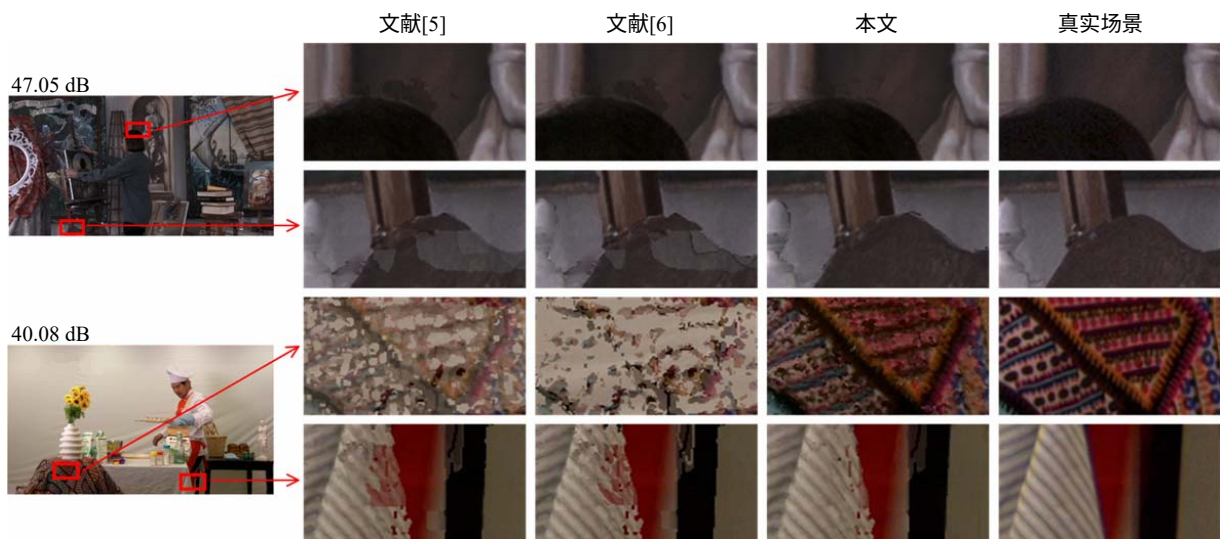


图10 主观质量对比

图, 可见, $TP_x_2y_1$ 视点的深度分布比较集中, 而 $EC3_x_2y_3$ 视点的深度分布相对分散, 同时前景纹理更加丰富, 深度图估计的不准确率和参考视点映射图像中的竞争区域都同时增大了。

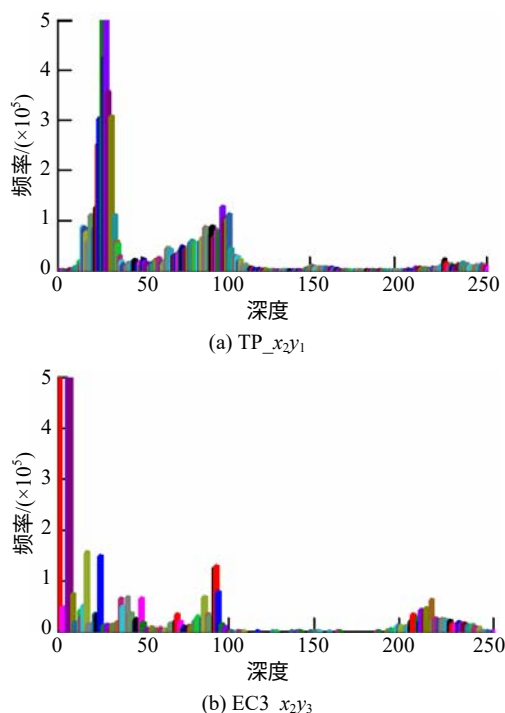


图 11 深度直方图

3.3 复杂度对比分析

本文实验硬件环境配置为 AMD Ryzen 5 3600 6-Core Processor@3.60 GHz, 16 GB 内存, 软件平台是 Visual Studio 2017. 表 4 统计了各算法占用的内存和绘制一帧虚拟图像消耗的平均时间. 本文算法的空间复杂度略高于对比算法, 是因为使用的测试序列分辨率较高, 使得每次需要处理的数据量较大. 对于每帧图像, 本文算法额外的消耗时间主要来源于自适应视点融合环节的 SSIM 计算, 占据了绘制过程 86.5% 的时间. 自适应视点融合的时间取决于各参考视点间竞争区域面积, 面积越大, 消耗时间越多. 但同时, 竞争区域面积越大, 本文算法效果越好.

表 4 复杂度对比

算法	占用内存/MB	消耗时间/s
文献[5]	308	2.25
文献[6]	309	2.38
本文	320	22.50

4 结论

针对平面相机阵列四参考视点的 DIBR 方案, 本文提出了一种深度和 SSIM 引导的四参考视点融合算法, 旨在消除虚拟视点图像绘制过程中竞争导致的伪影和背景渗透等失真. 本算法不完全依赖于参考视点的深度信息, 而是综合考虑了深度与图像块 SSIM 来优化各参考视点的融合权重, 使重建场景更接近于真实场景. 实验结果表明, 与对比算法相比, 本文算法可获得更高质量的虚拟视点绘制图像, 且各参考视点间竞争区域面积越大, 提升效果越明显. 不过, 虚拟视点图像质量的提升不可避免地增加了算法的时间成本, 在未来的研究中, 可结合时域信息进一步优化虚拟视点融合算法, 以提升用户的场景体验感.

参考文献:

- [1] Zhang Y, Zhu L, Hamzaoui R, et al. Highly efficient multiview depth coding based on histogram projection and allowable depth distortion[J]. IEEE Transactions on Image Processing, 2021, 30:402-417.
- [2] Teratani M, Senoh T, Kroon B, et al. Overview of MPEG-I visual test materials[C]//131st MPEG Meeting of ISO/IEC JTC1/SC29/WG11, 2020.
- [3] Gao X, Li K, Chen W, et al. Free viewpoint video synthesis based on DIBR[C]//2020 IEEE Conference on Multimedia Information Processing and Retrieval, 2020: 275-278.
- [4] Cheung C H, Sheng L, Ngan K N. Motion compensated virtual view synthesis using novel particle cell[J]. IEEE Transactions on Multimedia, 2021, 23:1908-1923.
- [5] Senoh T, Tetsutani N, Yasuda H, et al. Proposed View Synthesis Reference Software (pVSRS 4.3) manual[C]//124th MPEG Meeting of ISO/IEC JTC1/SC29/WG11, 2018.
- [6] Vijayanagar K R, Kim J, Lee Y, et al. Efficient view synthesis for multi-view video plus depth[C]//2013 IEEE International Conference on Image Processing, 2013: 2197-2201.
- [7] Lee T C, Chien C L, Hang H M. Virtual view synthesis quality refinement[C]//2016 IEEE 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2016:1-4.
- [8] Wegner K, Stankiewicz O, Domański M. Novel depth-

- based blending technique for improved virtual view synthesis[C]//2016 IEEE International Conference on Signals and Electronic Systems, 2016:93-98.
- [9] Sharma M, Ragavan G. A novel image fusion scheme for FTV view synthesis based on layered depth scene representation & scale periodic transform[C]//2019 IEEE International Conference on 3D Immersion, 2019:1-8.
- [10] Qiao Y, Jiao L, Yang S, et al. Color correction and depth-based hierarchical hole filling in free viewpoint generation[J]. IEEE Transactions on Broadcasting, 2019, 65(2):294-307.
- [11] 蔡李美, 李新福, 田学东. 基于分层图像融合的虚拟视点绘制算法[J]. 计算机工程, 2021, 47(4):204-210.
- [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] Liu A, Lin W, Paul M, et al. Just noticeable difference for images with decomposition model for separating edge and textured regions[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(11):1648-1652.
- [14] Wang X, Wang K, Yang B, et al. Perceptual quality assessment on DIBR synthesized videos with composite distortions[C]//2020 IEEE International Conference on Image Processing, 2020:186-190.
- [15] Arseneau S, Cooperstock J R. An improved representation of junctions through asymmetric tensor diffusion[C]// International Symposium on Visual Computing, 2006: 363-372.
- [16] Senoh T, Tetsutani N, Yasuda H, et al. Proposed Depth Estimation Reference Software (pDERS 8) manual[C]// 124th MPEG Meeting of ISO/IEC JTC1/SC29/WG11, 2018.
- [17] Dziembowski A. [MPEG-I Visual] Software manual of IV-PSNR for immersive video[C]//127th MPEG Meeting of ISO/IEC JTC1/SC29/WG11, 2019.

Depth and structural similarity guided blending algorithm with four-reference viewpoints

HE Xiaomei¹, ZHANG Lianjun¹, CHEN Fen^{1,2*}, CAI Zhenzhen², WANG Xiaodong¹

(1.Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China;

2.School of Electrical and Electronic Engineering, Chongqing University of Technology, Chongqing 400054, China)

Abstract: Depth image based rendering (DIBR) with four-reference viewpoints for planar camera array allows users to enjoy the scene immersively from all directions. This scheme can effectively eradicate holes distortion at the boundaries of virtual viewpoint images. However, it results in obvious artifacts, background penetration, etc. Therefore, this paper proposes a four-reference viewpoints blending algorithm guided by depth and structural similarity (SSIM). Firstly, the main distortions generated by DIBR are analyzed based on four-reference viewpoints for planar camera array comprehensively. Then, the view errors are eliminated as per the relative position relationship between reference and virtual viewpoints, and distortion masks of the blended image are extracted using the noticeable distortion model. In the end, the depth information and the SSIM in the distorted areas of each viewpoint are utilized to optimize the blending weights, and the high-quality virtual viewpoint image is thus rendered. Experimental results show that the SSIM and immersive video peak signal-to-noise ratio of the virtual viewpoint image rendered by the proposed algorithm reads from 0.001 8 to 1.46 dB, which is higher than those of the benchmark. In addition, the virtual viewpoint image rendered by the proposed algorithm is more consistent with the ground truth image compared with the state-of-the-arts in terms of visual perception.

Key words: depth image based rendering; viewpoints blending; plane camera array

(责任编辑 韩 超)