

可重构阵列处理器的虚拟视点合成算法设计

谢晓燕¹ 王昱¹ 武鑫² 朱筠²

¹(西安邮电大学计算机学院 陕西 西安 710121)

²(西安邮电大学电子工程学院 陕西 西安 710121)

摘要 针对三维高效视频编码(3D High Efficient Video Coding, 3D-HEVC)算法的计算量显著增加的问题,为满足高清视频(High Definition, HD)实时处理的要求,通过分析视点合成参考软件(View Synthesis Reference Software, VSRS)算法,提出一种重新配置架构。该架构可根据不同的场景向相应的处理单元发送相应的指令,采用均值滤波处理近景图像,中值滤波处理远景图像。仿真结果表明,该结构实现了测试图像的平均峰值信噪比为 34.55 dB,硬件设计工作在最大时钟频率为 160.2 MHz, Bee4 平台上的 VirTX-6 FF1759 LX550T FPGA 可实现输出每秒 124 帧的 720P(1 024 × 768) 视频。

关键词 3D-HEVC VSRS 深度图预处理 重新配置 可重构

中图分类号 TP3 TN919.71 **文献标志码** A **DOI**:10.3969/j.issn.1000-386x.2020.11.015

DESIGN OF VIRTUAL VIEW SYNTHESIS ALGORITHM FOR RECONFIGURABLE ARRAY PROCESSORS

Xie Xiaoyan¹ Wang Yu¹ Wu Xin² Zhu Yun²

¹(School of Computing, Xi'an University of Posts and Telecommunications, Xi'an 710121, Shaanxi, China)

²(School of Electronic Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, Shaanxi, China)

Abstract This paper proposes a reconfiguration structure by analyzing the viewpoint synthesis reference software (VSRS) algorithm, which reduces the computational complexity of the 3D efficient video coding (3D-HEVC) algorithm and meets the real-time requirements of high definition (HD) video. The structure could send instructions to the corresponding processing units according to different scenes. The close-range images were processed by mean filtering, and the distant view images was processed by median filtering. The simulation results show that the structure achieves an average Peak Signal to Noise Ratio(PSNR) of 34.55 dB for the test images. The hardware design works at a maximum clock frequency of 160.2 MHz. The Virtx-6 FF1759 LX550T FPGA on the Bee4 platform can output 720P (1 024 × 768) video at 124 frames per second.

Keywords 3D-HEVC VSRS Depth map preprocessing Reconfiguration Reconfigurable

0 引言

3D-HEVC 为多视点加深度(Multi-View plus Depth, MVD)的数据格式。基于深度图像渲染(Depth Image

Based Rendering, DIBR)算法^[1-3,19]中的两步映射,多视点的主要问题是映射中存在大量的数据传输和带宽的限制。三维视频编码团队 ISO/IEC MPEG 基于 DIBR 的两步映射开发了高质量视图合成参考软件 VSRS^[4],其中最重要的问题是合成虚拟视点中的

空洞^[5]。

视图合成技术遮挡产生的孔和通过扭曲产生的取整误差会导致渲染质量变差^[6]。深度图可以被深度相机捕获,但是由于深度相机太昂贵,因此当前深度估计成为获取深度图的主流方法,然而由深度估计算法计算的深度数据可能具有错误的值或空间不一致,这些问题可能导致所生成的图像上存在各种视觉伪影。可以通过深度预处理来减少孔,从而提高合成视图的质量。基于 Markov 随机有限元(Markov Random Fields, MRF)的快速视图合成方法^[7]保留了候选标签来保持合成视图中的视觉一致性,大大缩短了运行时间,但该候选约束缺少前景信息,使得视图质量有所下降。基于多损失卷积神经网络(Convolutional Neural Networks, CNN)的虚拟视点合成方法^[8]使用了基于三层网络的多损失函数,虽提高了合成视图的质量,但带来了极高的计算复杂度。

目前,减少空洞的普遍硬件实现方法分为三种:通用处理器(General Purpose Processor, GPP)实现、专用集成电路(Application Specific Integrated Circuit, ASIC)实现和可重构体系结构的实现。文献[9]提出了一种完整的包含视差估计和视差合成的立体视图 DIBR 算法,虽然该硬件体系结构有较高的处理速度,但其较高的成本及功率使其不具有普遍适用性。在超大规模集成电路(Very Large Scale Integration, VLSI)的实现中,文献[10]提出了一种基于相位的视点合成硬件系统,虽然该方法不受对象边界的影响,可以提高视图的质量,但是带来了较高的计算复杂度。文献[11]提出了一种新的虚拟视图合成方法,使用两个相邻的真实视图合成虚拟视图,但是其未修复区域由其他真实视图的信息填充,需要消耗更多的内存资源。文献[12]利用视图间纹理和深度图相似性,通过纹理优化(View Synthesis through Texture Optimization, VISTO)迭代生成一个虚拟视图,但因未使用深度图,使得合成视图没有立体感。

虚拟视点合成的 ASIC 实现虽然在一定程度上提高了操作效率,但其灵活性较差、成本更高。随着各种视频算法的计算和性能要求的提高,GPP 和 ASIC 的缺点日益突出。可重构体系结构逐渐成为研究热点,它将通用处理器的灵活性和专用硬件的效率相结合,这是一种高效的体系结构^[13]。文献[14]提出了一种对低复杂度的三维图像合成的可重构结构,在系统中动态管理不同的应用工作负载使其重新配置,缺点是对

复杂性较高的应用处理性能不高。在此背景下,本文提出基于可重构阵列处理器的实现方法,能够适应不同场景以满足图像质量和加快虚拟视图合成的重构结构。

1 VSRS 算法及并行分析

为了解决视觉伪影问题,采用三种滤波方式预处理深度数据,分别为:均值滤波、中值滤波和高斯滤波。用虚拟视图合成参考软件 VSRS3.5 对图 1 所示的测试序列 balloons、newspaper、kendo、lovebird1、poznstreet、dancer 和 GT-Fly 的深度图像进行不同程度的预处理,统计了无预处理、中值滤波、均值滤波和高斯滤波在不同窗口大小下 Y 分量各参考视点图像的 PSNR,统计结果如表 1 所示。具体过程如下:(1) 配置工程环境;(2) 设置为中值滤波,令其模板分别为(3,3)、(5,5)、(7,7)时,测出各测试序列的 PSNR 值,并记录;(3) 设置为均值滤波,令其模板分别为(4,4)、(5,5)、(6,6)测出各测试序列的 PSNR 值,并记录;(4) 设置为高斯滤波测出各测试序列的 PSNR 值,并记录。

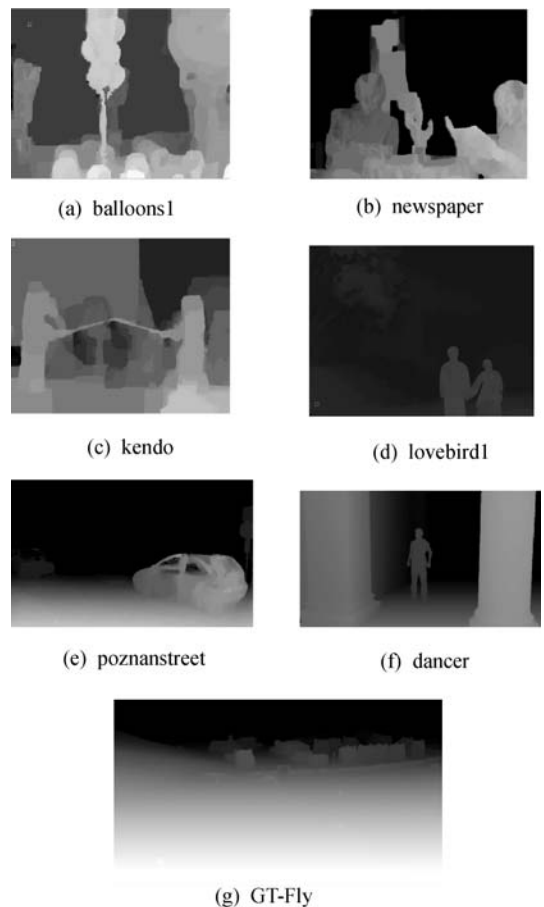


图 1 3D-HEVC 标准测试序列深度图

表1 3D-HEVC标准测试序列不同预处理下的PSNR统计结果

测试序列	处理前	中值滤波 (3,3)	中值滤波 (5,5)	中值滤波 (7,7)	均值滤波 (4,4)	均值滤波 (5,5)	均值滤波 (6,6)	高斯滤波 (5,5)
balloons1	38.4	38.7	38.6	38.6	38.7	38.7	38.7	38.8
newspaper	33.57	33.56	33.56	33.54	33.70	33.73	33.64	33.74
kendo	39.05	39.05	39.05	39.05	39.10	39.10	39.09	39.11
lovebird1	34.35	34.51	34.53	34.58	34.40	34.39	34.38	34.47
poznanstreet	38.89	39.04	39.10	39.15	38.81	38.85	38.70	38.87
dancer	38.72	38.76	38.77	38.80	38.75	38.73	38.731	38.73
GT-Fly	38.80	39.03	39.06	39.11	38.93	38.90	38.87	38.91
平均	37.397 71	37.522 00	37.524 86	37.547 14	37.484 86	37.486 14	37.444 43	37.518 00

可以看出,对于深度图 balloons1、newspaper、kendo,高斯滤波效果最好,对于深度图 lovebird1、poznanstreet、dancer、GT-Fly,中值滤波的图像质量更好。为了便于硬件实现,对于近景图像的深度图,本文采用与高斯滤波原理作用类似,模板为(5,5)的均值滤波进行预处理;对于远景图像的深度图,采用模板为(7,7)的中值滤波进行预处理。

1) 深度图预处理:表1中, balloons1、newspaper、kendo的测试结果表明当使用 5×5 模板时,均值滤波效果最好。将这25个数据两两相加后通过移位减的操作实现目标点的计算。lovebird1、poznanstreet、dancer、GT-Fly的测试结果表明模板为(7,7)用中值滤波效果最好。由于冒泡排序法对一组排序数据相关性要求较高,在对这49个数据进行排序时,为了得到更大的并行性,采用无数据相关性的双调排序进行排序。

2) 3D-Warping:本文采用更有利于实时显示高清视频的1D模式来完成3D-Warping。为了提高算法执行效率,在计算摄像机模型的特征矩阵时采用查找表(Look-Up-Table, LUT)的方式对相同深度值进行统一计算,避免了重复操作。在VSRS软件操作中3D-Warping完成后才进行空洞填充操作,这样串行的设计相当耗时,为此可将像素点映射变化与空洞填充并行化处理,如图2所示。

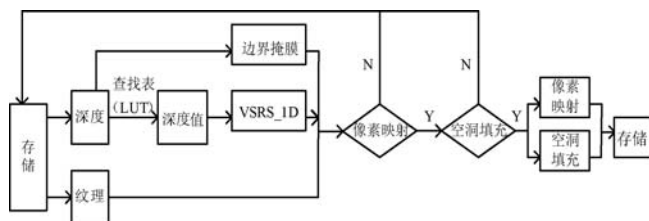


图2 像素映射与空洞填充的并行化

3) 空洞填充:在虚拟视点合成软件测试模型VSRS3.5中,通过将邻近点背景像素沿着水平方向传播至空洞的方式进行填充。首先通过掩膜后的合成视点判断空洞的位置,接着判断该空洞点的下一个位置

是否为空,直到判断完整个水平方向。若为空则标记空洞像素点加1,否则判断该像素点的背景是否在右侧:若在右侧,则使用右边的点进行填充,否则使用左边的点进行填充,如图3所示。整个填充过程中的重点在于引入深度数据区分背景像素是否在图像的右边。对于 1024×768 的图像来说至少需要786432次循环才能完成,这有碍于高清视频实时显示。对于整幅半高清图像来说,填充时每行的数据是不相关的,因此可以采用并行化实现。

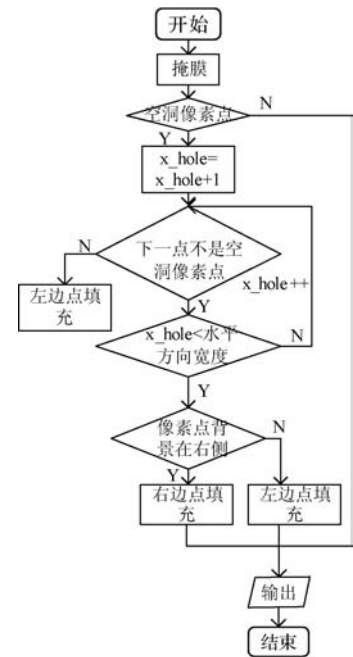


图3 VSRS中空洞填充算法流程图

2 算法实现

2.1 可重构阵列处理器

图4为实现VSRS的硬件体系结构。其中,合并的图像存储在外部存储器中。处理单元(Processing Element, PE)通过数据输入存储器(Data Input Memory, DIM)将外部存储器中的数据加载到本地数据存储器中,

计算掩模的类型,并将其反馈到全局控制器。全局控制器比较接收的掩模类型数据并将相应的指令发送到 PE 的指令存储器。

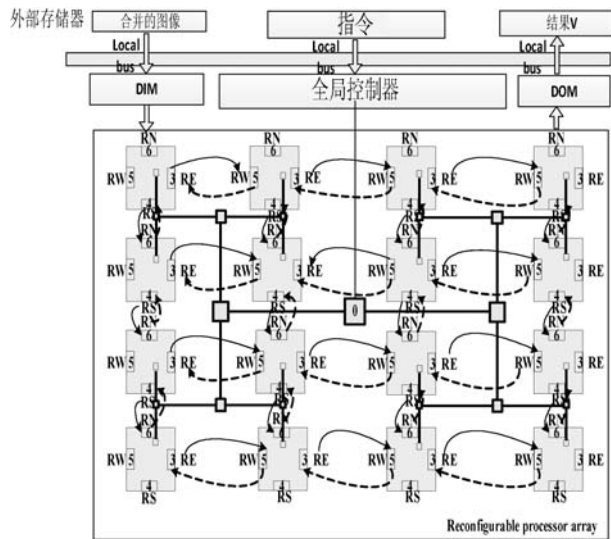


图 4 VSRS 算法的硬件体系结构

全局控制器用于控制和管理阵列计算资源,包括操作指令的广播、调用指令的分配和计算资源信息的收集。上层是指令存储器,下层是由 4×4 个 PE 组成的阵列处理器。主要功能是在主机接口与阵列处理器之间形成一个层次化的编程网络,并利用层次化编程网络实现数组计算资源的控制和管理。为了得到简单的寻址,寻址过程的比特宽度可以逐级递减,并且确保每个指令同时到达 PE。分层编程网络设计为通过 H 网络加载指令。

为了节省硬件资源,本文利用视频阵列处理器的动态可重构机制来切换深度图预处理算法。其核心是在指令下发送网络,主要使用指令网络中的指令广播操作和指令发送操作,指令广播操作负责执行 PE 操作。该方法是在数组结构中的每个 PE 的指令存储器中存储指令,然后通过指令广播操作打开全部或部分 PE。指令发送操作主要负责执行模式的重新配置,通过发布指令将指令从全局指令存储发送到指定 PE。

可编程可重构阵列处理器 (Dynamic Reconfigurable Array Processor, DPR-CODEC) 是由项目组自主研发的一款动态视频阵列处理器。本文基于 DPR-CODEC 来设计 VSRS 算法,DPR-CODEC 的部分结构如图 5 所示。阵列处理器由 32×32 个 PE 以邻接互连形式组成(图 5 中仅以 4×4 结构来示例),采用物理不分层、逻辑分层的划分方式。由于视频算法中数据的处理基本都是以 $N \times N$ 的矩形块进行,且可重构阵列处理器能够适应不同场景,所以这种体系结构相比其他结构更加满足视频算法需求,能够更有效地对视频算法进行并行化设计。

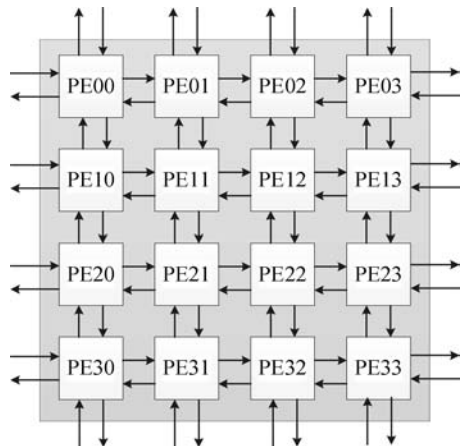


图 5 DPR-CODEC 的结构

2.2 并行化方法

VSRS 算法分为深度图预处理、边界掩膜、3D-Warping 和空洞填充,通过设计 4 个 PEG (Processing Element Group, PEG) (每个算法用一个 PEG) 来完成整体映射,如图 6 所示。具体映射步骤如下:左视图深度图 (Deep Left, DL)、右视图数据 (Deep Right, DR) 是两个参考视图深度图,将其存储在外部存储中。首先,PEG10 和 PEG00 同时从外存获取深度图数据,进行边界掩膜和深度图预处理操作。PEG10 执行完毕后通过 PE1000、PE1001、PE1002、PE1003 邻接互连传给 PEG00,执行完毕后通过 PEG00 簇的 PE03、PE13、PE23、PE33 邻接互连传递给 PEG01。接着 PEG01 根据边界掩膜信息和预处理后的图像进行像素点映射变换后产生的左视点深度图 (Deep virtual Left, DVL) 和右视点深度图 (Deep virtual Right, DVR) 存入外存。最后 PEG02 从外存加载变形后的深度图进行空洞填充,并将结果 V 存入外存。

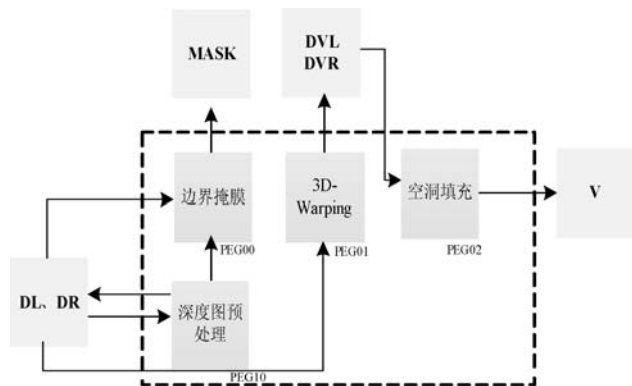


图 6 可重构虚拟视点合成算法映射方案

3 实验

对测试模型 VSRS3.5 的配置文件进行修改,得到测试数据;通过 b4d_bram_write 函数将测试数据下发至 FPGA 中的 left0、left1、left2、left3 共享寄存器中;通

过 b4d_bram_read 函数读取共享寄存器 result 中的数据,并将采集到的数据存储到片外存储器中,接着将并程序的指令初始化为指令存储器;用 Questasim-64—10.1d 对可重构阵列的结构进行验证。在并行化方案的功能仿真中,对 Bee4 芯片 VIETEX-6FF1759 LX550T FPGA^[17]集成。实验表明,整个 PE 的工作频率为 159 MHz,片上资源利用率为 57 280 个 LUT 和 84 850 个 Slice。

表 2 是在 LoveBird1 测试序列下实现虚拟视图合成算法的视频阵列结构并行实现的周期统计。可重构阵列处理器兼顾专用硬件的高效性和通用处理器的灵活性,本文算法虽然占用了大量的资源,但工作频率更接近专用硬件,其灵活性比专用硬件高。此外,本文采用可重构阵列实现了中值滤波和均值滤波的预处理算法,以适应不同的视频场景。FPGA 测试表明,对于分辨率为 1 024 × 768 的测试序列 LoveBird1,处理速度可以达到 124 fps, Y-PSNR 为 34.55。

表 2 虚拟视点合成算法的执行周期和时间统计

VSRS 算法	周期 (256 data)	720P	时间/s	吞吐量
深度图 预处理 (中值滤波)	214	657 408	4.21×10^{-3}	720P@237 fps
深度图 预处理 (均值滤波)	409	1 256 448	8.05×10^{-3}	720P@124 fps
边界掩膜 Warping 空洞填充	176 280 362	540 672 860 160 1 112 060	3.47×10^{-3} 5.51×10^{-3} 7.13×10^{-3}	720P@289 fps 720P@181 fps 720P@140 fps

表 3 比较了 10 帧的 PSNR 的平均合成性能,使用捕获的虚拟视图和合成的视图来计算合成的 PSNR。综合性能表明,本文算法的差分分别为 0.043 dB 和 0.145 dB。

表 3 10 帧平均 PSNR 的比较

测试序列	相机号 L, V, R	Original VSRS 3 5/dB	本文算法 /dB
balloons	1, 3, 5	38.43	38.65
newspaper	2, 3, 4	33.55	33.71
kendo	3, 4, 5	39.00	39.05
lovebird1	4, 5, 6	34.47	34.53
poznastreet	3, 4, 5	39.03	39.21
dancer	1, 5, 9	38.76	38.81
GT-Fly	9, 5, 1	38.87	39.16

表 4 为在视频阵列体系结构中并行虚拟视点合成算法的性能比较。文献[15]虽然带宽减少,但图像质量较差,在 LoviBrd1 的相同测试序列下,PSNR 在图像质量上提高了 2.748 dB,吞吐量提高了 40%。在文献[17]中提出的实时自由视图合成硬件使用了由 607K LUT 和 303K 寄存器组成的 VixTeX-XC7VX85T FPGA,以实现 175 MHz 的工作频率,吞吐量为 720P@ 55 fps。本文使用一个由 124K LUT 和一个 39K 寄存器组成的 VITEX-7 XC7VX85T FPGA,以达到 191 MHz 的工作频率,吞吐量为 720P@ 152 fps。与文献[17]相比,本文提出的可重构阵列结构提高了 176% 的吞吐量,减少了 82% 的资源。提出的立体视图多视图转换系统在文献[18]中的硬件设计可以在 160.2 MHz 的最大时钟频率下操作,实时地通过 HDMI 接口每秒 60 帧输出 1 080P 视频,文献[18]使用由 40.2K LUT 和 87.7K 寄存器组成的 Stulx-3 FPGA,以实现 160 MHz 的工作频率。与文献[18]相比,本文提出的可重构阵列结构减少了 14% 的吞吐量,增加了 27% 的资源,虽然吞吐量略低于文献[18],但图像质量 PSNR 提高了 4.35 dB。此外,本文提出的架构优点在于:可以根据不同的视频场景切换深度图预处理操作,以实现更高质量的合成视图,具有更高的灵活性,更适合于市场应用。

表 4 不同硬件结构的性能比较

结构	芯片	最大工作频率 /MHz	Gate/LUT 数量	速率 /fps	PSNR /dB	吞吐量
文献[15]	90 nm	200	268.5K Gate	88	31.80	720P@ 88 fps
文献[16]	65 nm	160	4 424K Gate	30	—	1 080P@ 30 fps
文献[17]	Virtex-7	175	607 000 LUTs 303 000 Registers	55	—	720P@ 55 fps
文献[18]	Stratix-3	160	40 200 LUTs 88 700 Registers	66	30.20	1 080P@ 66 fps
本文	Virtex-6	156	124 848 LUTs 39 248 Registers	47	34.55	1 080P@ 47 fps
本文	Virtex-7	191	124 848 LUTs 39 248 Registers	152	34.55	720P@ 152 fps

4 结 语

为了兼顾 3D-HEVC 中 VSRS 算法处理速度与图像质量两个指标,提出了一种基于可重构阵列处理器结构的重构方案,可用于近景预处理和远景预处理之间的切换。为了实现虚拟视图合成算法,通过对 3D-HEVC 官网提供的七个典型序列进行测试,结果表明该体系结构可以达到实时处理的要求,并可以正确地切换预处理算法。与文献[15]相比,该方案在相同资源占用的情况下,提高了 40% 左右的速度,PSNR 提高了 2.748 dB。与文献[16]相比,该方案提高了 176% 左右的速度,减少了 82% 的资源。虽然吞吐量略低于文献[18],但图像质量 PSNR 提高了 4.35 dB。最突出的优点是采用可重构结构,方便了不同用户的使用,具有较好的灵活性,更适合市场应用的推广。同时,空洞填充是虚拟视点合成极为重要的部分,针对空洞填充部分的优化,有待下一步改进,使得最终虚拟视点合成视频的质量更佳。

参 考 文 献

- [1] Horng Y R, Tseng Y C, Chang T S. Stereoscopic image generation with directional Gaussian filter [C] // Proceedings of IEEE International Symposium on Circuits and Systems. IEEE, 2010: 2650 – 2653.
- [2] Chen W Y, Chang Y L, Lin S F, et al. Efficient depth image based rendering with edge dependent depth filter and interpolation [C] // 2005 IEEE International Conference on Multimedia and Expo. IEEE, 2005: 1314 – 1317.
- [3] Park Y K, Jung K, Oh Y, et al. Depth-image-based rendering for 3DTV service over T-DMB [J]. Image Communication, 2009, 24(1/2): 122 – 136.
- [4] Wegner K, Tech G, Yea S, et al. 3D-HEVC draft text 4 [C] // Joint Collaborative Team on 3D Video Coding Extension Development, 2014.
- [5] Cho J H, Song W, Choi H, et al. Hole filling method for depth image based rendering based on boundary decision [J]. IEEE Signal Processing Letters, 2017, 24(3): 329 – 333.
- [6] Rahaman D M M, Paul M. Virtual view synthesis for free viewpoint video and multiview video compression using gaussian mixture modelling [J]. IEEE Transactions on Image Processing, 2018, 27(3): 1190 – 1201.
- [7] Luo G B, Zhu Y S, Guo B. Fast MRF-based hole filling for view synthesis [J]. IEEE Signal Processing Letters, 2018, 25(1): 75 – 79.
- [8] Nian Z C, Jung C. High-quality virtual view synthesis for light field cameras using multi-loss convolutional neural networks [C] // 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018: 2605 – 2609.
- [9] Li Y Z, Claesen L, Huang K. A real-time high-quality complete system for depth image-based rendering on FPGA [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(4): 1179 – 1193.
- [10] Huang H C, Wang Y C, Chen W C, et al. System and VLSI implementation of phase-based view synthesis [C] // 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019: 1428 – 1432.
- [11] Dziembowski A, Grzelka A, Mieloch D, et al. Multiview synthesis—Improved view synthesis for virtual navigation [C] // 2016 Picture Coding Symposium (PCS). IEEE, 2016: 1 – 5.
- [12] Sun W X, Au O C, Xu L F, et al. Seamless view synthesis through texture optimization [J]. IEEE Transactions on Image Processing, 2014, 23(1): 342 – 355.
- [13] Li T, Xiao L Z, Huang H C, et al. PAAG: A polymorphic array architecture for graphics and image processing [C] // Fifth International Symposium on Parallel Architectures, Algorithms and Programming. IEEE, 2012: 242 – 249.
- [14] Amor N B, Lahbib K, Frikha T. Design of a dynamically reconfigurable architecture for the 3D image synthesis [C] // 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). IEEE, 2017: 1 – 5.
- [15] Horng Y R, Tseng Y C, Chang T S. VLSI architecture for real-time HD1080p view synthesis engine [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2011, 21(9): 1329 – 1340.
- [16] Schaffner M, Gurkaynak F K, Greisen P, et al. Hybrid ASIC/FPGA system for fully automatic stereo-to-multiview conversion using IDW [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2016, 26(11): 2093 – 2108.
- [17] Akin A, Capocchia R, Narinx J, et al. Real-time free viewpoint synthesis using three-camera disparity estimation hardware [C] // IEEE International Symposium on Circuits and Systems. IEEE, 2015: 1908 – 1908.
- [18] Kuo P C, Lo K L, Tseng H K, et al. Stereoview to multiview conversion architecture for auto-stereoscopic 3D displays [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2017, 27(99): 1.
- [19] Chen Y, Tech G, Wegner K, et al. Test model 11 of 3D-HEVC and MV-HEVC [C] // Joint Collaborative Team on 3D Video Coding Extension Development, 2015.