7.1 A 216fps 4096×2160p 3DTV Set-Top Box SoC for Free-Viewpoint 3DTV Applications

Pei-Kuei Tsung, Ping-Chih Lin, Kuan-Yu Chen, Tzu-Der Chuang, Hsin-Jung Yang, Shao-Yi Chien, Li-Fu Ding, Wei-Yin Chen, Chih-Chi Cheng, Tung-Chien Chen, Liang-Gee Chen

National Taiwan University, Taipei, Taiwan

3DTV promises to become the mainstream of next-generation TV systems. High-resolution 3DTV provides users with a vivid watching experience. Moreover, free-viewpoint view synthesis (FVVS) extends the common two-view stereo 3D vision into virtual reality by generating unlimited views from any desired viewpoint. In the next-generation 3DTV systems, the set-top box (STB) SoC requires both a high-definition (HD) multiview video-coding (MVC) decoder to reconstruct the real camera-captured scenes and a free-viewpoint view synthesizer to generate the virtual scenes [1-2].

There are three main challenges to design an efficient high-resolution 3DTV STB SoC: (1) High processing capability of both the real-view decoder and virtual-view synthesizer is required to support various 2D/3D applications in HDTV. For example, 66.2TOPS of computation is consumed to synthesize one virtual view in quad full-HD (QFHD) at 30fps. (2) In order to support FVVS, including 3D translation and 3D rotation (6D), matrix-based warping is needed for every pixel. In this case, two adjacent pixels in the reference view may be warped to the arbitrary positions in the virtual view according to their depth and epipolar geometry. Horizontal raster-scan-based scheduling [4-5] cannot deal with these irregular pixel relationships and can only support 1D horizontal shifts in the view synthesis. (3) The block-based memory accessing of reference pixels is not suitable for FVVS because of the processing nature of the irregular pixel access. 31.5GB/s system memory bandwidth is thus required for each virtual view in QFHD. State-of-the-art 3DTV chips cannot solve the issues above [4-6].

Our 3DTV STB SoC is summarized as follows. First, a hardware-oriented 6D FVVS flow is introduced along with the corresponding architecture to solve the first two design challenges. A maximum 1911MPixel/s throughput is achieved, and is 9-to-40.5× higher than the previous works [4-6]. Second, the cache-based texture reorder architecture with the dynamic warping reference frame selection (DWRFS) scheme reduces the external memory bandwidth by 95.7% in view synthesis. Finally, the precision-optimized Homographic Transform (HT) and the single-iteration inpainting save 68% area in the warping engine and 93.3% of computing cycles, respectively.

Figure 7.1.1 shows the 6D FVVS flow and the target applications. The MVC decoder reconstructs the real-view videos, the corresponding camera matrix, and the depth values from the bitstream. The view synthesizer then generates the virtual views from the real views. As shown in the left of Fig. 7.1.1, the pixels on one specific epipolar line of the virtual-view frame can only be found along the corresponding epipolar line in the reference frame. Therefore, the processing schedule along the epipolar lines avoids the conflict of memory accessing compared with the traditional horizontal raster-scan schedule. To support the various epipolar geometries, 7 kinds of block patterns are used to access reference pixels with slopes between ±45°. The accessed pixels are then reordered into an 8×8 block in the texture-reorder stage. For the slopes larger than ±45°, rotating the scan order of blocks converts the effective slope within ±45° and thus extends the supporting rotation angle to ±180°. The warping stage is capable of performing the accurate geometry transformation to further support the continuous epipolar geometry from the 7 discrete slopes. After the warping, the occlusion regions on the virtual-view blocks are then filled in the inpainting stage. The results are outputted after inverse reordering in the final stage.

Figure 7.1.2 shows the system architecture. The first 3 stages are the MVC decoder while the remaining 4 stages constitute the 6D FVVS flow. The parallel mode is shown at the bottom of Fig. 7.1.2. Since the textures in one virtual view are the subset of the neighboring virtual view in the 3D-translation-only cases, one reference block loaded from the system bus can be reused to generate multiple virtual views in parallel. In this way, the throughput is boosted, and the data access bandwidth is reduced. Furthermore, a full-utilization mode is designed by further increasing the parallelism of views and reusing the decoder bus to increase the output bandwidth. The throughput of 216fps corresponding to 9 views @ 24fps is achieved for QFHD, and is 12.5-to-40.5× higher than the previous works [4-5].

Figure 7.1.3 shows the bandwidth-reduction schemes. In the software-based algorithm [3], multiple reference views are jointly utilized in the warping to generate one virtual view. The textures provided by these reference views have large overlapped regions. In order to avoid the bandwidth to access the redundant information, the DWRFS sets one reference view as the main reference, and others are loaded only for the occlusion regions. Furthermore, the block-based pixel accesses on the system bus conflict with the epipolar-based processing pattern in the texture reordering and inverse reordering stages. A texture-reordering cache and a 16×24 inverse reordering buffer are designed to efficiently load the reference pixels and write back the results in the bus burst mode as shown on the left and right sides of Fig. 7.1.3. The system memory bandwidth for view synthesis is reduced by 95.7%.

Figure 7.1.4 shows the warping and inpainting engines. The warping engine requires 256 matrices corresponding to 256 depth values in HT. To save the large memory requirements from these 256 matrices, a linear interpolation (LI) scheme is used. Only 3 warping matrices are saved in LI. For other depth values, matrices are linearly interpolated from these 3 matrices. As a result, area is reduced by 68%. The inpainting engine is shown on the bottom-right of Fig. 7.1.4. In the software-based inpainting algorithms [3], the occluded pixels are padded from the neighboring pixels by iteratively calculating and updating the gradient of the textures. In the proposed inpainting algorithm, the pixels are filled in a single iteration by analyzing the cause of occlusion of each pixel. The inpainting process can thus be parallelized, and the cycle count is reduced by 93.3%.

Figure 7.1.5 shows the detailed chip specifications and the example FVVS results. The core size is 5.76mm^2 including 1416K logic gates and 19.9KB on-chip SRAM in TSMC 40nm CMOS. 6D FVVS functionality, H.264/AVC High-Profile decoding, and MVC High-Profile decoding are supported in a single chip. The maximum FVVS capability is 4096×2160p 216fps (24fps @ 9 views), and the maximum decoding capability is 4096×2160p 30fps.

The performance evaluation is shown in Fig. 7.1.6. Compared with the previous works, the chip supports 12.5-to-40.5× view synthesis, 1.25-to-26.7× decoding, and 9-to-40.5× overall system capabilities. Furthermore, 6D FVVS is supported in addition to the conventional 1D horizontal shift. The total power efficiency of the chip is 27.5MPixel/mW, which is 6.6-to-229× higher than those of the previous works. The chip micrograph is shown in Fig. 7.1.7.

Acknowledgement:
The authors thank TSMC University Shuttle Program and Morly Hsieh for process support and National Chip Implementation Center for chip testing. This work is funded by National Science Council and TSMC.

References:
[1] Joint Video Team of ISO/IEC MPEG & ITU-T VCEG, "Joint Draft 8.0 on Multiview Video Coding," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, July, 2008.
[2] MPEG-FTV Group, "Draft Report on Experimental Framework for 3D Video Coding" ISO/IEC JTC1/SC29/WG11 MPEG2010/N11273, April. 2010.
[3] MPEG-FTV Group, "Reference Softwares for Depth Estimation and View Synthesis" ISO/IEC JTC1/SC29/WG11 M15377, April. 2008.
[4] S. H. Kim, et al., "A 36fps SXGA 3D Display Processor with a Programmable 3D Graphics Rendering Engine," *ISSCC Dig. Tech. Papers*, pp. 276–277, Feb. 2007.
[5] S. H. Kim, et al., "A 116fps 74mW Mobile Heterogeneous 3D-Media Processor for 3D Display Contents," *Symposium on VLSI Circuits*, pp. 258-259, June 2009.
[6] T. D. Chuang, et al., "A 59.5mW scalable/multi-view video decoder chip for Quad/3D Full HDTV and video streaming applications," *ISSCC Dig. Tech. Papers*, pp. 330–331, Feb. 2010.

 978-1-61284-302-5/11/$26.00

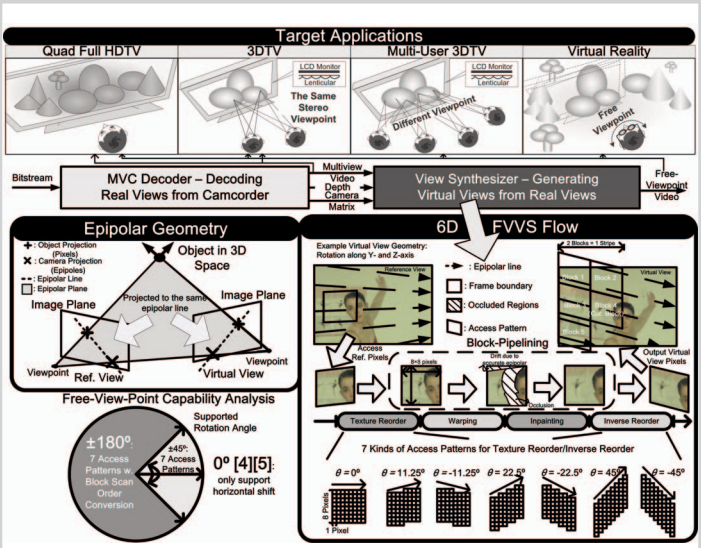

Figure 7.1.1: Virtual view generation flow and the target applications.

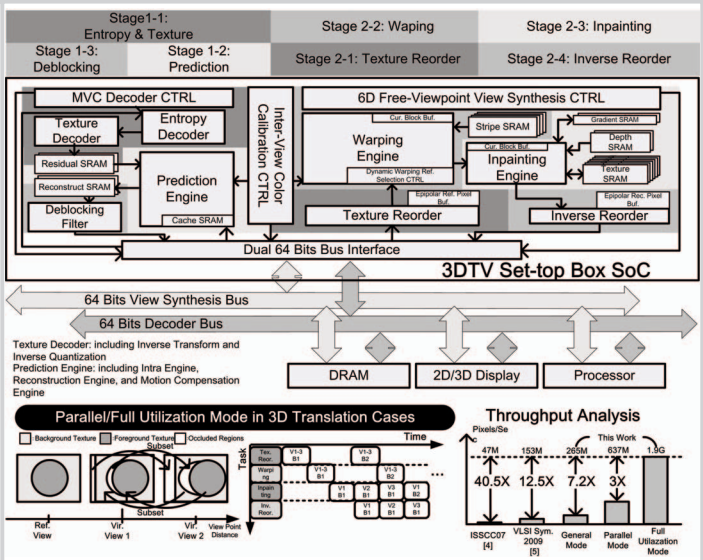

Figure 7.1.2: System architecture of the 3DTV set-top box SoC.

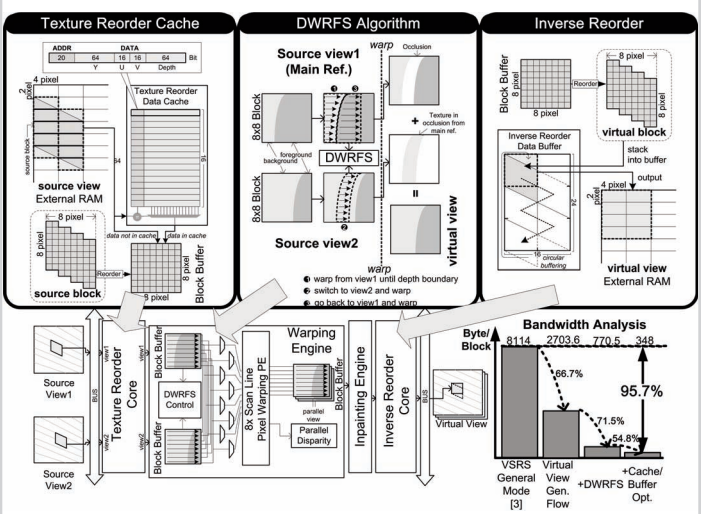

Figure 7.1.3: Bandwidth-reduction schemes.

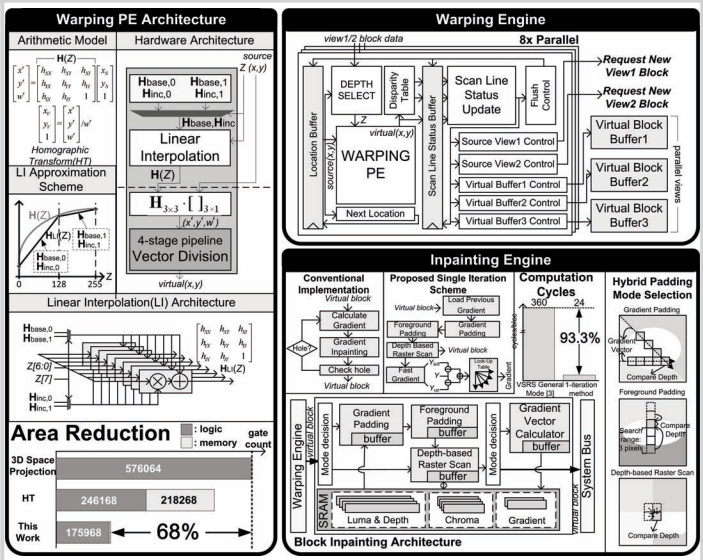

Figure 7.1.4: Architecture of the warping and inpainting engine.

Technology	TSMC 40nm 1P10M CMOS
Supply Voltage	Core: 0.9V; I/O: 2.5V
Temperature	25°C
Die Size	3.1×3.1mm²
Core Size	2.4×2.4mm²
Logic Gate Count	1416K (2-input NAND gate)
On-Chip SRAM	19.9 KB
Free-View-Point 3D Features	Full Functionality 3DTV Free-View-Point Geometry Supporting (3- dimensional translation & 3-dimensional rotation)
Decoding Features	H.264/AVC High Profile H.264/AVC MVC High Profile

3DTV Set-top Box Functionalities

Mode Name	Maximum Virtual View Synthesis Throughput	Maximum Real View Decoding Throughput	Power Consumption
Full Utilization	2160p 216fps (24fps@9Views)	Not Supported	69.5mw@240MHz
Hybrid Parallel	2160p 72fps (24fps@3Views)	2160p 30fps	81.6mw@240MHz
Hybrid General	2160p 30fps	2160p 30fps	81.5mw@240MHz
Parallel Mode	2160p 72fps (24fps@3Views)	Not Supported	68.2mw@240MHz
General Mode	2160p 30fps	Not Supported	71.4mw@240MHz
Decoder Mode	Not Supported	2160p 30fps	51.4mw@240MHz

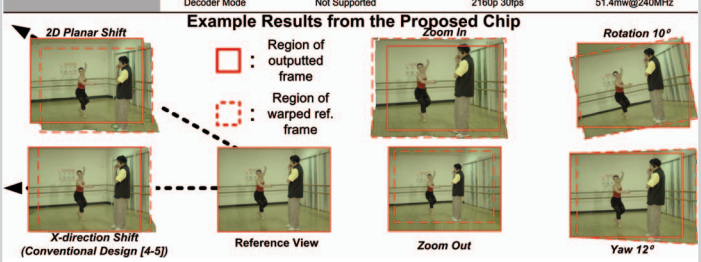

Figure 7.1.5: Chip features and different configurations.

	ISSCC 2007 [4]	VLSI Sym. 2009 [5]	ISSCC 2010 [6]	This Work
Max View Synthesis Capability	1280×1024@36fps	1280×1024@116fps	Not Supported	4096×2160@216fps
Free-Viewpoint 3D Functionality *	1 Dimension	1 Dimension	Not Supported	6 Dimensions
Max Decoder Capability	Not Supported	352×288×2view @49fps	4096×2160@24fps	4096×2160@30fps
Supported Standard	Not Supported	N/A (Only MC and color space conversion part)	H.264 High Profile MVC High Profile SVC High Profile	H.264 High Profile MVC High Profile
Power Consumption	379mW@50MHz	36mW@60MHz	59mW@200MHz	69.5mW@240MHz
Total Power Efficiency **	0.12MPixels/mW	4.2MPixels/mW	3.6MPixels/mW	27.5MPixels/mW
Technology	0.13um	0.13um	90nm	40nm
Logic Gate Count	1744K	930K	414K	1416K
On-Chip SRAM	N/A	N/A	9.0KB	19.9KB

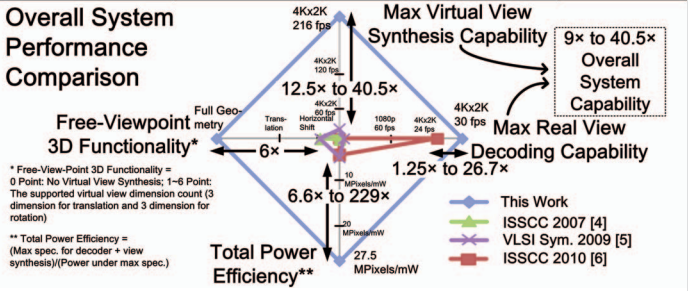

Figure 7.1.6: Comparison with the state-of-the-art 3DTV chips.

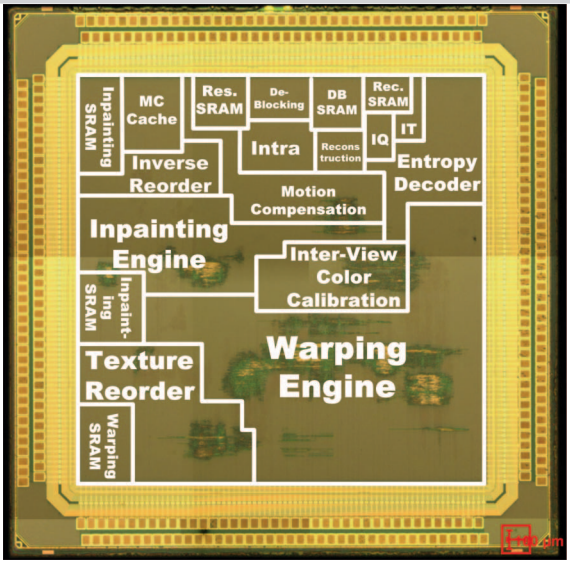

Figure 7.1.7: Chip micrograph.

978-1-61284-302-5/11/$26.00