9.6 Reconfigurable Processor for Energy-Scalable Computational Photography

Rahul Rithe[1], Priyanka Raina[1], Nathan Ickes[1], Srikanth V Tenneti[2], Anantha P Chandrakasan[1]

[1]Massachusetts Institute of Technology, Cambridge, MA, [2]California Institute of Technology, Pasadena, CA

Computational photography applications such as lightfield photography [1] enable capture and synthesis of images that could not be captured with a traditional camera. Non-linear filtering techniques like bilateral filtering [2] form a significant part of computational photography. These techniques have a wide range of applications, including High-Dynamic Range (HDR) imaging [3], Low-Light Enhanced (LLE) imaging [4], tone management and video enhancement. The high computational complexity of such multimedia processing applications necessitates fast hardware implementations [5] to enable real-time processing. This paper describes a hardware implementation of a reconfigurable multi-application processor for computational photography.

A software-based bilateral grid structure is described in [6], which enables fast bilateral filtering but requires a large amount of storage (65MB for a 10Mpixel image) for processing. In this work, we implement bilateral filtering using a reconfigurable grid, which reduces the storage requirement to 21.5kB by scheduling the filtering engine so that only two grid rows need to be stored at a time. The implementation is flexible to allow varying grid sizes for energy/resolution scalable image processing. The reconfigurable filtering engine performs HDR imaging, LLE imaging and glare reduction, as shown in Fig. 9.6.1. The filtering engine can also be accessed from off-chip and used with other applications. The implementation significantly accelerates bilateral filtering and enables various edge-aware image processing applications in real-time on HD images. The testchip is able to process a 10Mpixel image in 771ms with 17.8mW power consumption while operating at 98MHz, 0.9V.

The bilateral grid structure used by this chip is constructed as follows. The input image is partitioned into blocks of size $\sigma_s \times \sigma_s$ and a histogram of pixel intensity values is generated for each block. Each histogram has $256/\sigma_r$ bins. This results in a 3D representation of the 2D image, referred to as the bilateral grid where each grid cell (i, j, r) stores the number of pixels in a block corresponding to that intensity bin (W^r_{ij}) and their summed intensity (I^r_{ij}). The grid assignment (GA) engine, shown in Fig. 9.6.2, performs this operation. The convolution (Conv) engine convolves the grid intensities and weights with a 3×3×3 Gaussian kernel, which is equivalent to bilateral filtering in the image domain [6], and returns the normalized intensity. The interpolation engine reconstructs the filtered 2D image from the filtered grid. The filtered intensity value at pixel (x, y) is obtained by trilinear interpolation of a 2×2×2 filtered grid values surrounding the location (x/σ_s, y/σ_s, I_{xy}/σ_r). To meet throughput requirements, the interpolation engine is implemented as three pipelined stages of linear interpolations.

The grid processing tasks are scheduled to minimize local storage requirements and memory traffic. Fig. 9.6.3 shows the architecture of the bilateral filtering engine and task scheduling. Grid processing is performed cell-by-cell in a row-wise manner. When cell (i, j) is being assigned, the convolution engine is processing cell (i-2, j-1) and the interpolation engine is processing cell (i-4, j-2). Boundary rows and columns are replicated for processing boundary cells. This scheduling scheme allows processing without storing the entire grid. Only two grid rows need to be stored locally at a time. The number of grid cells varies inversely with σ_s and σ_r. Most applications work well with a coarse grid resolution on the order of 32 pixels. Decreasing the number of grid cells directly reduces the number of computations required. The grid size is configurable by adjusting σ_s from 16 to 128 and σ_r from 16 to 64. For a 10Mpixel (4096×2592) image, the number of grid cells scales from 663552 (σ_s = 16, σ_r = 16) to 2592 (σ_s = 128, σ_r = 64). The 21.5kB of on-chip SRAM is used to store two rows of created and filtered grid cells. The SRAM is implemented as 8 banks supporting a maximum of 256 cells in each row of the grid with 16 intensity levels, corresponding to the worst case of σ_s = 16, σ_r = 16. Each bank is clock and input gated to save energy when a lower resolution grid is used. Only 1 bank is used when σ_s = 128 and all 8 banks are used when σ_s = 16.

The testchip contains two bilateral filter engines, each processing 4 pixels/cycle. Fig. 9.6.4 shows the architecture of the HDR creation module. It takes one low-dynamic range (LDR) pixel each from 3 different exposures (I_{E1}, I_{E2}, I_{E3}) and merges them into an HDR pixel (I_{HDR}) using camera response curves. Displaying HDR images on LDR media requires tone mapping that compresses image dynamic range by non-linear filtering. A tone-mapped HDR image (I_{TM}) is created by bilateral filtering HDR intensity values in the log domain followed by contrast reduction [3]. In HDR mode, both bilateral grids are configured to perform filtering in an interleaved manner, where each grid processes alternate blocks in parallel. Glare reduction is similar to performing single image tone mapping and is integrated with the HDR architecture. LLE imaging is performed by merging two images captured in quick succession, one taken without flash (I_{NF}) and one with flash (I_F). The bilateral grid is used to decompose both images into base and detail layers. In this mode, one grid is configured to perform bilateral filtering on the non-flash image and the other to perform cross-bilateral filtering [6] on the flash image using the non-flash image. The scene ambience is captured in the base layer of I_{NF} and details are captured in the detail layer of I_F. The flash image contains shadows that are not present in the non-flash image. A novel shadow correction module, shown in Fig. 9.6.4, is implemented which merges the details from the flash image with base layer of the cross-bilateral filtered non-flash image and corrects for the flash shadows to avoid artifacts. A mask representing regions with high detail in the filtered non-flash image is created and details from the flash image are added in the masked regions only. The processing is done in 4×4 sub blocks from $\sigma_s \times \sigma_s$ blocks to reduce complexity. This implementation of the shadow correction module handles shadows effectively to produce LLE images without artifacts.

The testchip is implemented in 40nm CMOS technology and verified to be operational from 25MHz at 0.5V to 98MHz at 0.9V. Fig. 9.6.5 shows outputs for HDR imaging, LLE imaging and glare reduction. This chip is designed to function as an accelerator core as part of a larger microprocessor system, utilizing the system's existing DRAM resources. For standalone testing of this chip a 32b wide 266MHz DDR2 memory controller was implemented using a Xilinx XC5VLX50 FPGA. The energy vs. performance trade-off and the frequency of operation of the testchip is shown in Fig. 9.6.6 for a range of V_{DD}, along with runtimes for different image sizes at 98MHz with 0.9V V_{DD}. The runtime for a 10Mpixel image is compared with GPU/CPU implementations of C++ code that replicates the functionality of the testchip. The processor achieves 15× reduction in run-time compared to the CPU implementation, while consuming 17.8mW of power, an energy reduction of at least three orders of magnitude compared to previous CPU or GPU implementations [6]. The architecture supports a high amount of parallelism, which can be used to further enhance the throughput and reduce the runtime. The energy scalable implementation proposed in this work enables efficient integration into portable multimedia devices for real-time computational photography.

Acknowledgements:
This work was funded by Foxconn Technology Group. The authors thank TSMC University Shuttle Program for chip fabrication and Prof. Fredo Durand and Jonathan Regan-Kelley for valuable feedback and suggestions.

References:
[1] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, P. Hanrahan, "Light-field photography with a handheld plenoptic camera," *Stanford University Computer Science Tech Report CSTR 2005-02*, April 2005.
[2] C. Tomasi, R. Manduchi, "Bilateral Filtering for Gray and Color Images," *IEEE International Conf. on Computer Vision*, pp. 839-846, 1998.
[3] F. Durand, J. Dorsey, "Fast Bilateral Filtering for the display of high-dynamic-range images," *ACM Trans. on Graphics*, vol. 21, no. 3, pp. 257-266, 2002.
[4] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, K. Toyama, "Digital Photography with flash and no-flash image pairs," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 664-672, 2004.
[5] J. Oh, G. Kim, J. Park, I. Hong, H.-J. Yoo, "A 320mW 342GOPS Real-Time Moving Object Recognition Processor for HD 720p Video Streams," *ISSCC Dig. Tech. Papers,* 220-221, 2012.
[6] J. Chen, S. Paris, F. Durand, "Real time edge-aware image processing with the bilateral grid," *ACM Transactions on Graphics*, vol. 26, no. 3, article 103, 2007.

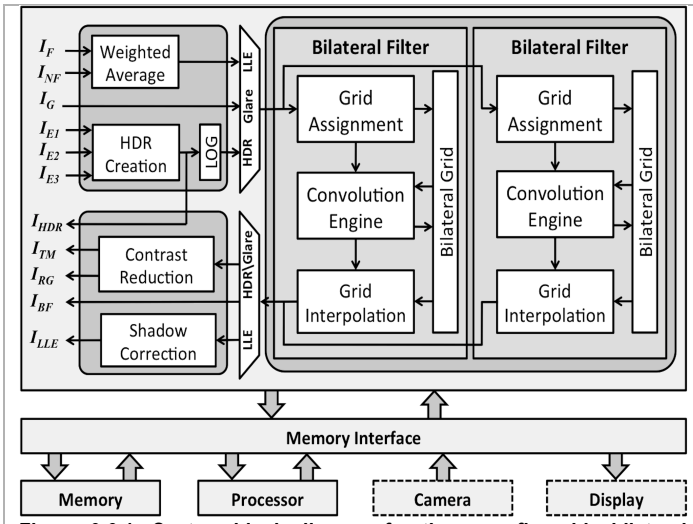


Figure 9.6.1: System block diagram for the reconfigurable bilateral filter engines.

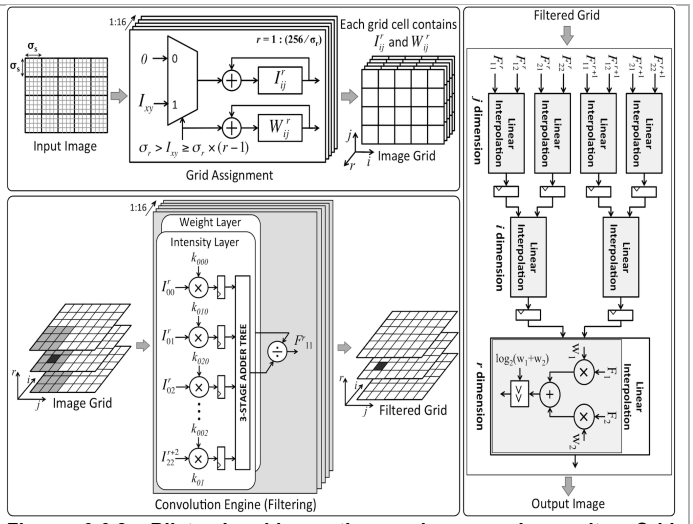


Figure 9.6.2: Bilateral grid creation and processing units: Grid assignment engine, convolution engine and interpolation engine.

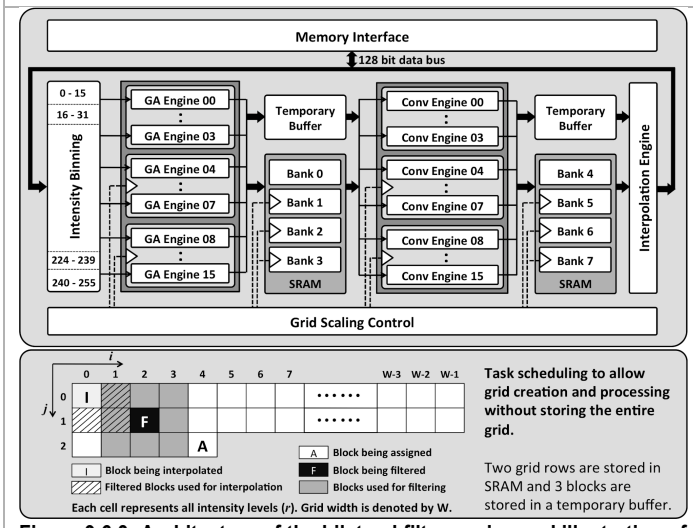


Figure 9.6.3: Architecture of the bilateral filter engine and illustration of task scheduling. Grid scalability is achieved by gating processing engines and SRAM banks.

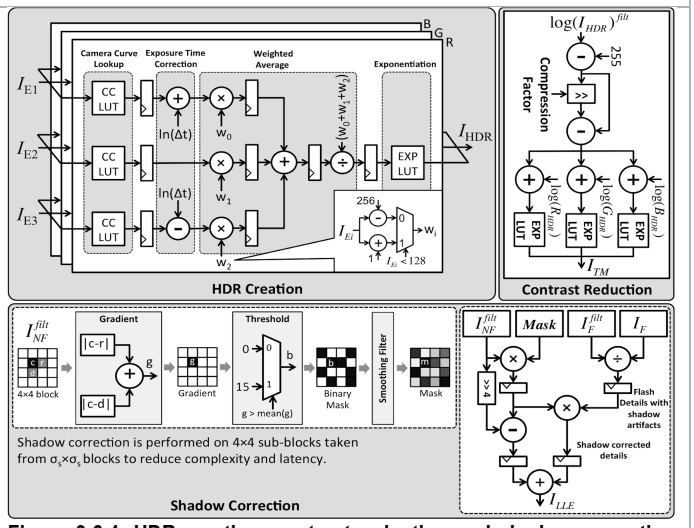


Figure 9.6.4: HDR creation, contrast reduction and shadow correction modules.

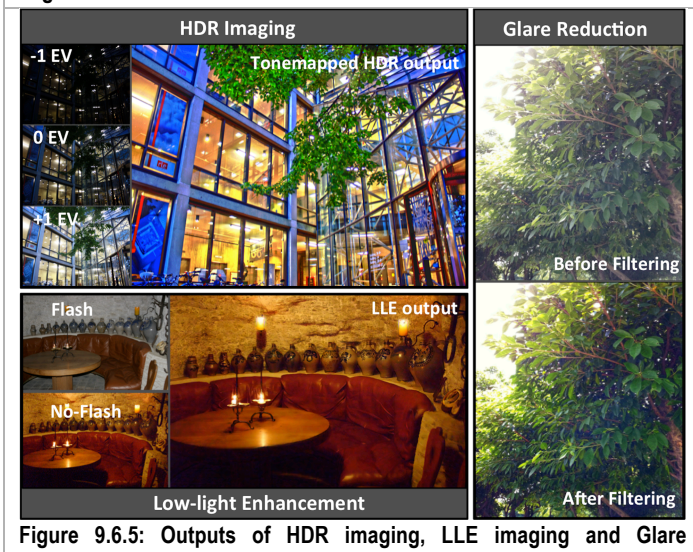
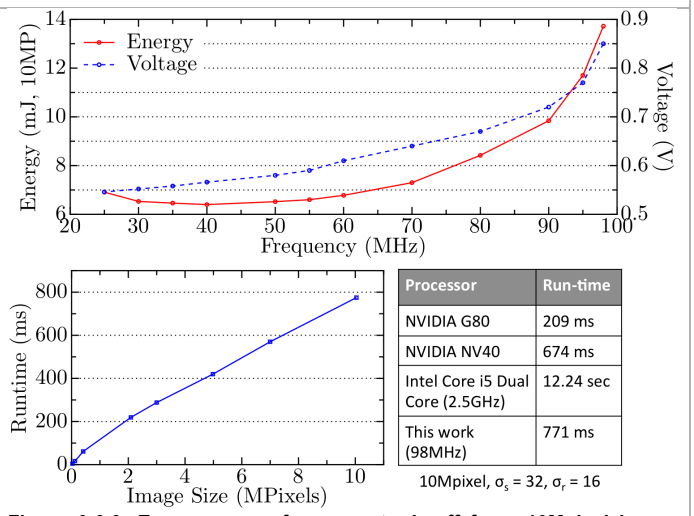


Figure 9.6.5: Outputs of HDR imaging, LLE imaging and Glare reduction.



Processor	Run-time
NVIDIA G80	209 ms
NVIDIA NV40	674 ms
Intel Core i5 Dual Core (2.5GHz)	12.24 sec
This work (98MHz)	771 ms

10Mpixel, σ_s = 32, σ_r = 16

Figure 9.6.6: Energy vs. performance trade-off for a 10Mpixel image and the frequency of operation for a range of voltages. Run times for different image sizes at 98MHz, 0.9V.