Video image signal processing (ISP) has come a long way from the time of analog signals. Today, digital signal processing makes possible image data manipulation at the bit level, offering unprecedented control over image quality. Digital signal processing, of course, is not the same as a digital signal processor, or DSP. While DSP has been popular for video image signal processing in the digital domain, ISP can be implemented by a number of processing devices: DSPs, ASICs, ASSPs and, increasingly, field programmable gate arrays, or FPGAs.

There are several reasons for the growing popularity of FPGAs. Two of those reasons reflect recent trends in security cameras that dramatically increase the quantity of image data that needs to be processed, and the third is an economic reason – the cost of the camera bill of materials (BOM). There are two major trends that are changing the way security cameras are architected: the advent of megapixel sensors, and the need for high (or wide) dynamic range (HDR/WDR).

There was a time when a VGA resolution sensor was sufficient for security camera purposes, usually viewed by an operator or simply archived for later review. However, with the dramatic increase in the number of security cameras used worldwide, there are not enough human operators available and so the security industry has begun to rely on software to analyze the video, either in real time or later, to discover if anything untoward occurred in the region of interest. Sophisticated video analytics (VA) algorithms have been developed to highlight the extraordinary from the ordinary; however, in order to be effective, these algorithms need much more detail that can be provided by VGA resolution cameras. Cameras need higher resolution for VA to be able to discern general movement in restricted and/or large areas, e.g. a parking lot is at capacity. A camera needs approximately 30 pixels/in for license plate recognition, and approximately 150 pixels/in to view more detailed activity, such as identifying cash register transactions. One megapixel covers the detail in a seven foot by seven foot area, and it would take four VGA cameras to match the power of a one megapixel camera. Image sensors have been developed, and are commercially available, for one, two, five and even ten megapixel resolution. Obviously, as the number of pixels has increased, so too has the amount of data that must be processed to take advantage of the increased resolution.

High dynamic range (HDR) also known as wide dynamic range (WDR), measures how well the sensor and the ISP function see into both dark and brightly lit areas. We are all familiar with amateur outdoor family pictures taken with the sun behind the people in the photograph. While the landscape bathed in sunlight is bright and clear, people faces are quite dark. This happens because the (usually automatic) camera adjusts its exposure to the bright sunlight in the scene. That exposure, however, is too short to properly register the darker objects. If one manually sets the exposure or aperture to let in more light, one will be able to discern detail in the dark areas, but at the expense of detail in the bright areas, which now are overexposed and completely washed out.

This is not a good result for either human operators or for VA software, since much of the detail in the region of interest is lost. HDR sensors solve this problem in creative ways, all of which depend on capturing multiple images, each with different exposure times, and then having the ISP pipeline combine and blend these images to preserve and render visible detail from both bright and dark areas in the region of interest. Obviously, multiple exposures for the same image translate into an increased amount of data to be processed. For example, when a video camera that outputs full HD 1080p images at 60 frames per second is working with a HDR sensor that takes 3 exposures per frame, the ISP pipeline inside the camera is actually processing the equivalent of 60 x 3,
or 180, frames per second. A megapixel sensor and HDR combine to dramatically increase the processing load of the ISP pipeline. DSP devices, being sequential engines by nature, struggle to keep up with this tremendous data processing load. It may still be possible to process the data in our example of a 1080p60 HDR pipeline in high-end DSPs, but with cost and power consumption that is prohibitive and commercially unviable. FPGAs, due to their inherent parallelism, are ideally suited to take on the increased load of high definition, high dynamic range image signal processing. In addition to providing high performance at very low power and cost, FPGAs are by definition programmable, which offers significant advantages over ASICs and ASSPs. ASICs are extremely expensive to design and build and, once built, cannot be altered. ASSP-based camera designs can be feature-limited by what is already baked into the standard parts, which are impossible to modify. In fact, several DSP and other ASSP devices in the video image signal processing market need an FPGA bridge between the sensor and the standard part in order to accommodate new serial interfaces that sensor manufacturers are using in order to get megapixel data off their sensors. With an FPGA-based implementation, camera manufacturers can take advantage of programmability to quickly adapt their designs to new sensors and technologies, or rapidly modify their ISP algorithms. In order to implement ISP with HDR in an FPGA, one must implement, at a minimum, the ISP blocks in the image signal processing pipeline shown in figure 1. The following ISP blocks are required:

- **Sensor Port, with auto black level correction:** this is required to detect and configure the image sensor registers and capture image data.
- **Automatic exposure:** the purpose of the automatic exposure block is to constantly adjust exposure to adapt to changing light conditions in real time.
- **Linearization:** the Aptina MT9M024/34 HDR sensor, for example, outputs 20 bits of information per color channel. In order to minimize the number of physical lines coming out of the sensor, Aptina uses a clever compression scheme to compress this data to 12 bits. Linearization is the process of decompressing this 12 bit data to recover the original 20 bits. Defective pixel correction: dead or hot pixels present in the sensor due to manufacturing processes are corrected with the defective pixel correction block. This block corrects the defective pixel with interpolated values based on neighboring pixels of the same color channel. Typical correction methods include detection of cold or hot pixels using either median or averaging estimation on the immediate pixel neighborhood.

2-D Noise Reduction: apart from cold and hot pixels, sensor pixels can randomly be noisy from frame to frame. This means that they output either too much or too little intensity in comparison with neighboring pixels. 2D noise reduction corrects for noisy pixels with interpolated values based on neighboring pixels of the same color channel. The resulting image is a mosaic of the three colors. To obtain a full-color image, various de-mosaic algorithms are used to interpolate a set of complete red, green, and blue values for each pixel.

Color correction matrix (CCM): image sensors often provide incorrect color rendition due to so-called cross-color effects that result from signal cross-talk between pixels. This effect leads to wrong color images (e.g. green with too much blue). Color correction involves complex matrix multiplication of pixel data to achieve clean colors.
Automatic white balance (AWB): sensors are not good at recognizing colors. AWB adjusts other colors in an image with reference to an inferred white color in the image through a so-called grey world algorithm. AWB determines white by examining the frequency (therefore wavelength) of incoming light, and renders the image with natural colors.

Gamma correction: sensor pixels react to the intensity of incoming light in a linear way. In order to provide pixel data to common video systems, such as a CRT tube with logarithmic response, conversion to a non-linear value encoding may be needed. Gamma correction provides this conversion.

High / wide dynamic range (HDR/WDR) processing: this is the block responsible for mapping 20 pixels of sensor data to 8 bit RGB in a way that renders both bright and dark areas of the image visible in the rendered image. A wide internal pipeline is required to ensure that no detail in dark areas is lost even when, for example, an intruder shines a flashlight directly into the camera lens. HDR, working in close conjunction with a fast-auto-exposure algorithm, can rapidly adjust exposure in changing light conditions.

Table 1 shows the typical FPGA resources used for implementing all the above ISP blocks in a 33K look-up-table (KLUT), low—cost, low—power, FPGA. In addition to the ISP blocks already mentioned, the actual implementation data incorporates a statistics engine that generates image histograms used by specific blocks in the system, a Lattice Mico32 soft processor for dynamic pipeline control, an I2C master to control various signals, a HDMI PHY block to drive HDMI signals directly off the FPGA and even a graphics overlay of a logo. This demonstrates that it is possible to fit an entire image signal processing pipeline, plus HDMI output, inside a low-cost, low-power FPGA such as the Lattice ECP3-35. The internal HDR pipeline is 32-bits wide, resulting in the ability to provide 192dB (20 log 2**32) of high dynamic range. In this real world implementation, a sensor with 120dB dynamic range was used, limiting the HDR to 120dB – still the highest of any FPGA implementation available. The actual implementation is capable of processing 1080p images at 60 frames per second while providing 120dB of HDR. As shown, a simple low cost 33KLUT FPGA easily handles a 1080p60 pipeline. The BOM for a 1080p60 HDR camera implemented with a Lattice ECP3-35 consists primarily of the sensor, the FPGA and associated clock oscillator, resistors and capacitors, voltage regulator, a HDMI connector and lens assembly. The implementation shown offers 120dB of HDR, 1080p60 performance, the fastest auto-exposure and very high quality auto-white balance in the industry. The LatticeECP3 is significantly lower in static and dynamic power consumption than competing FPGAs or DSPs. The FPGA supports the use of DDR3. Manufacturers wishing to incorporate frame buffer memory into their designs can take advantage of this capability to utilize high-performance, low-cost DDR3 memory in their camera designs. A low-power SERDES-capable FPGA enables manufacturers to implement a HDMI PHY directly inside the FPGA, providing HDMI functionality without the added cost of an external HDMI chip.