High Dynamic Range Video for Photometric Measurement of Illumination

Jonas Unger, Stefan Gustavson, VITA, Linköping University, Sweden[1]

ABSTRACT

We describe the design and implementation of a high dynamic range (HDR) imaging system capable of capturing RGB color images with a dynamic range of 10,000,000 : 1 at 25 frames per second. We use a highly programmable camera unit with high throughput A/D conversion, data processing and data output.

HDR acquisition is performed by multiple exposures in a continuous rolling shutter progression over the sensor. All the different exposures for one particular row of pixels are acquired head to tail within the frame time, which means that the time disparity between exposures is minimal, the entire frame time can be used for light integration and the longest exposure is almost the entire frame time. The system is highly configurable, and trade-offs are possible between dynamic range, precision, number of exposures, image resolution and frame rate.

Keywords: high dynamic range imaging, HDR, multiple exposures, video, rolling shutter

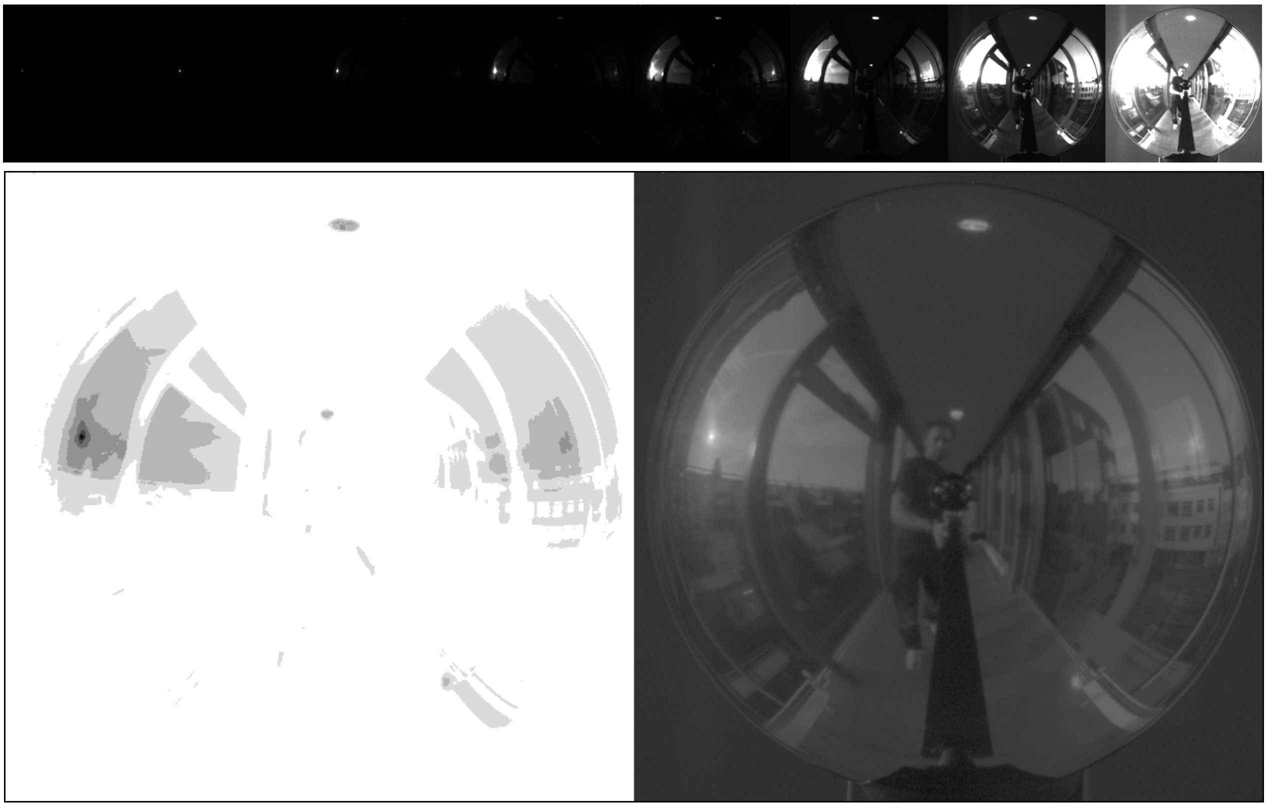


Figure 1: Sample image from a video stream from the system. Top: eight separate 8-bit linear exposures. Bottom left: index map showing which exposure was selected for each pixel in the final composited HDR image. Bottom right: log-mapped HDR image. The resolution is 512×512 pixels, and the full HDR image was captured in 40 ms. The dynamic range of this image exceeds 1,000,000:1, and all pixels in the image have a valid photometric value.

1. jonun@itn.liu.se, stegu@itn.liu.se

1. INTRODUCTION

In recent years, digital camera technology has undergone a rapid improvement in quality and a dramatic price drop, to the point where digital cameras are starting to replace film-based solutions for most applications. However, in terms of photometric range and precision, current digital camera technology does not compare favorably to traditional, film-based camera systems. To some extent this limitation is due to the acquisition process, but a large part of the problem has been the inherent limitation of most image file formats to 8 bits of precision per channel. We will refer to such technology as low dynamic range (LDR) imaging systems. The dynamic range is mostly extended somewhat by a nonlinear response curve implemented in software in the camera. Some special purpose camera systems extend the photometric precision of the output data somewhat, to 10 or 12 bits or sometimes more, but such systems are not common.

1.1. HDR image data

Recently, so-called high dynamic range (HDR) imaging has come into widespread use in computer graphics [1], both as an output format for synthetic imagery and as input data to image based lighting methods in the form of *light probes*, panoramic images of the incident light at a single point. With the introduction of HDR images into the production pipeline many old conventions in computer graphics have been modified or abandoned, and the image quality has improved to a level where it can be superior to traditional imaging methods. The field of computer graphics is moving to a more solid and physically motivated foundation based on photometry, instead of the ad hoc intensity values and crude reflectance models which were used in the past. This makes synthetic images share many important traits with images of the real world, and it also makes it possible to use results from physics and real world measurement data directly for photorealistic rendering. Useful file formats for HDR data have been defined and become de facto standard, although these formats still lack some important features, like efficient, high quality compression schemes. Support for HDR image file formats for data input and output is now implemented in most commercial renderers, albeit not always with completely seamless integration. Nevertheless, HDR images are definitely here to stay, because floating point data with a wide dynamic range is clearly a better data format than 8-bit integers for representing arbitrary light intensities.

1.2. HDR imaging methods

A number of useful HDR image file formats have been defined and agreed upon by the computer graphics industry, so HDR data input and output in renderers is a straightforward software implementation issue. However, there are still problems with HDR image acquisition from the real world. It is cumbersome and slow, the process is dependent on special purpose post-processing software to assemble the images, and the photometric calibration is also somewhat of a problem.

Most current HDR imaging methods are based on multiple exposure techniques as introduced by Madden [2] and made more widely known by Mann and Picard [3], where a sequence of LDR images are acquired using different exposure settings, and a final HDR image is assembled from several source images. Using this technique and a traditional LDR camera, it is difficult to acquire HDR images at high speed. The data for one single HDR image requires at least several seconds to capture and process, and both the camera and the scene need to be stationary during that time.

Some recent HDR imaging methods are instead based on logarithmic sensors. The dynamic range of a logarithmic sensor is impressive, and the speed is not a problem becuse only a single exposure is required. However, as reported by e.g. Krawczyk [4], logarithmic sensors still have relatively low image quality and SNR ratio, partly due to the relative immaturity of the sensor technology, but mainly due to the fact that the entire dynamic range needs to be represented by one single A/D converted value. Alternate and hybrid methods for dynamic range extension like the one presented by Nayar et al [5] is an increasingly popular research topic but, as of yet, no single method has emerged that solves all problems.

Among the HDR capture systems which have been presented previously, only logarithmic sensors can be said to do true HDR capture at video speed. Multiple exposure HDR video capture methods have been suggested by Waese and Debevec [6], although with a very low resolution, and by Kang et al [7], although with a low frame rate, considerable time disparity problems and an unsatisfactory dynamic range.

1.3. Our contribution

The system presented here uses a traditional CMOS sensor and a very fast multiple exposure technique for high quality, high speed HDR image capture. It is a significant extension and improvement over previous work which was performed in our lab and presented in 2003 [8]. The system is based on an existing smart sensor architecture with considerable on-chip high speed parallel processing and control logic. The capture is implemented as a rolling shutter progression for maximum efficiency. Resulting images are of very high quality, have a very wide dynamic range and can be captured, processed and saved at video resolution and video frame rates. To the best of our knowledge, our prototype system outperforms other existing technologies and is currently unique, but it is assembled from standard, off-the-shelf components, and the HDR algorithm is implemented entirely in software, partly in the camera and partly in the host computer. The system is being used for our own research in image based lighting [9], but its applications extend into any other areas of imaging where either true HDR video or rapid spatially resolved photometry could be useful.

2. HARDWARE PLATFORM

2.1. Camera

The hardware platform is a commercial smart sensor camera, the Ranger C55 from the company SICK IVP in Sweden[1]. The main intended use for this camera is laser range imaging by triangulation, but the on-chip processing logic of the sensor is general enough to allow many different imaging and image processing algorithms to be implemented. The sensor resolution is 1536 by 512 pixels. For each of the 1536 columns there is an A/D converter, a programmable bit-serial processing element (PE) and some local memory. The PEs operate in a strict SIMD fashion, taking their instruction feed from a common sequencer, and there is a simple general-purpose CPU within the camera, with some local memory and an operating system to handle data communication and some other tasks. The sensor architecture is presented in detail in [10].

The data output from the camera is a high-speed CameraLink digital video interface. A more recent version of the camera has a gigabit Ethernet connection instead, which performs about equally well but is simpler to connect to standard computer equipment.

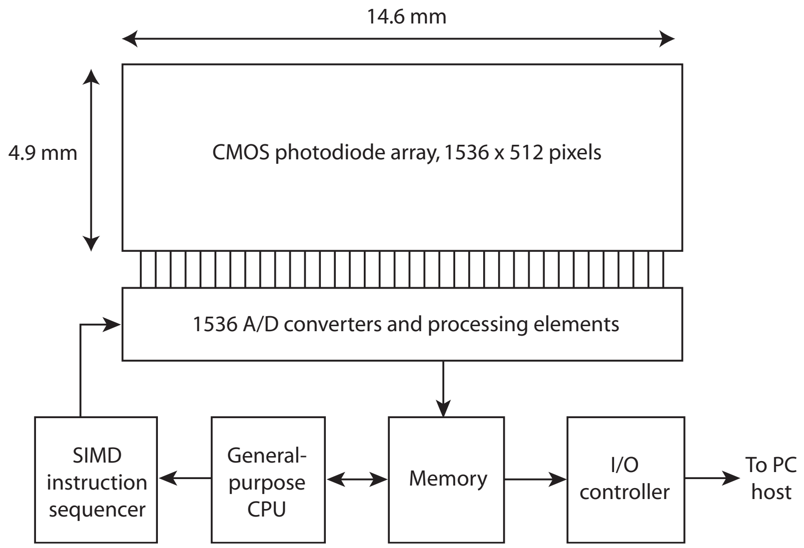
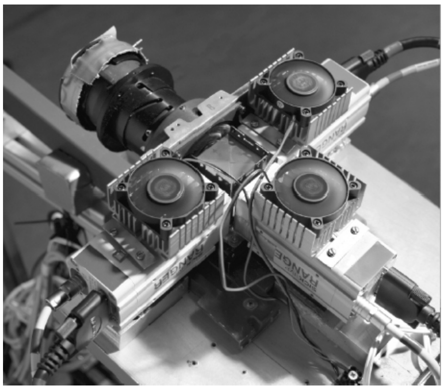


Figure 2: System overview of the Ranger C55 camera hardware (simplified)

1. http://www.sickivp.se

2.2. Color filters

The sensor is monochrome. In order to acquire color images, we have used two methods: one three-chip system fitted with dichroic beamsplitter optics to separate a single image into three spectral bands suitable for RGB imaging, and one with three separate cameras mounted side by side, each with identical lenses but different color filters. The beamsplitter with its single viewpoint properties would of course be the preferred method for general imaging applications, but we experienced practical problems with the calibration and robustness of that optical system, and there were some issues with polarization. Because our main application is image based lighting, we are only interested in a panoramic view of the camera surroundings, and this is achieved by imaging a reflective sphere. Three separate cameras pointing towards the same reflective sphere from some distance away will yield almost the same field of view from almost the same vantage point, and that is enough for our purposes. The color channels are adjusted to correspondence by a simple non-uniform resampling step in software.



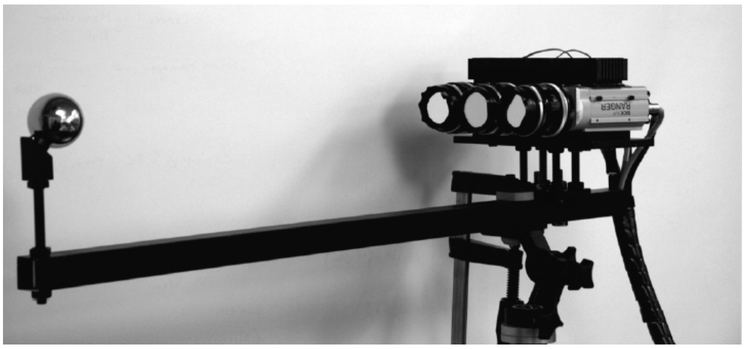


Figure 3: The single lens setup with an RGB beamsplitter (left), and the simpler setup with three separate lenses and RGB color filters pointing at a reflective sphere (right). Both use three separate, unmodified but reprogrammed C55 camera units.

3. CAPTURE ALGORITHM

The camera comes with a large library of standard software, aimed at industrial inspection and general image processing. The camera developers at SICK IVP kindly allowed us NDA access to their development software, and we have reprogrammed the camera at the microcode level to perform HDR capture. The result is a file with machine level code which can be run in any C55 camera unit, without any access to the development libraries.

3.1. Principle

Continuous multiple exposure capture at video frame rate is achieved by a software-controlled rolling shutter progression over the sensor, where each row is reset, read out and A/D converted in rapid succession several times during a frame. Not all rows will be exposed simultaneously, so the aperture needs to be fixed and only the exposure time can be varied. However, this is not a severe limitation, as the sensor can be programmed for exposure times as short as a single microsecond. After deciding on a frame rate, in our case 25 frames per second, the available frame time, 40 milliseconds, is split between the different exposure times, where the longest exposure is allowed a duration of almost the entire frame time. The exposure time is only a passive wait, which can be entirely hidden in other operations performed in parallel, like A/D conversion and data transfer. For an implementation with 512 rows of output, each row has $40\text{ms}/512 = 78\mu\text{s}$ for A/D conversion and data output, which is precisely enough for 8 exposures.

3.2. Details

The 1536 parallel A/D converters are simple 8-bit linear ramp converters, and one 8-bit conversion by itself takes 9.6 µs. The gain setting of the A/D conversion amplifiers can be varied and, in order to gain some extra precision in the digital values from the low end of the scale, we convert some analog values twice, once with unity gain and once with 4x gain. This effectively adds two extra bits of precision to low values, making the A/D converted value comparable in quality to a

10-bit value. The thermal noise is of course also amplified by the gain, but by cooling the camera housing to room temperature with a heat sink and fan assembly, the thermal noise of the sensor can be kept below 1 LSB for all but the longest exposure. Cooling to below room temperature would reduce the thermal noise even further.

The dynamic changes in gain increase the possible step size between exposure times. The actual exposures can be as far apart as 4 f-stops (16x) while still keeping the quantization error to less than 2 %. We always have at least 6 significant bits of precision for the lowest A/D converted values which need to be used for every exposure, except for the longest one where we use all but the very lowest values. Because the exposures are 4 f-stops apart, the longest exposure time is almost the entire frame time, and the available time is well utilised for light integration. Exposure times for our sample implementation are given in Table 1. Our current application uses 8 images from five different exposure times ranging from 10 µs to 37 ms, with a 4x gain extension to the three longest exposures, 8-bit A/D conversion, a frame rate of 25 FPS and an image resolution of 512 lines of up to 896 pixels each. A simplified timing diagram of the capture algorithm is given in Figure 4.

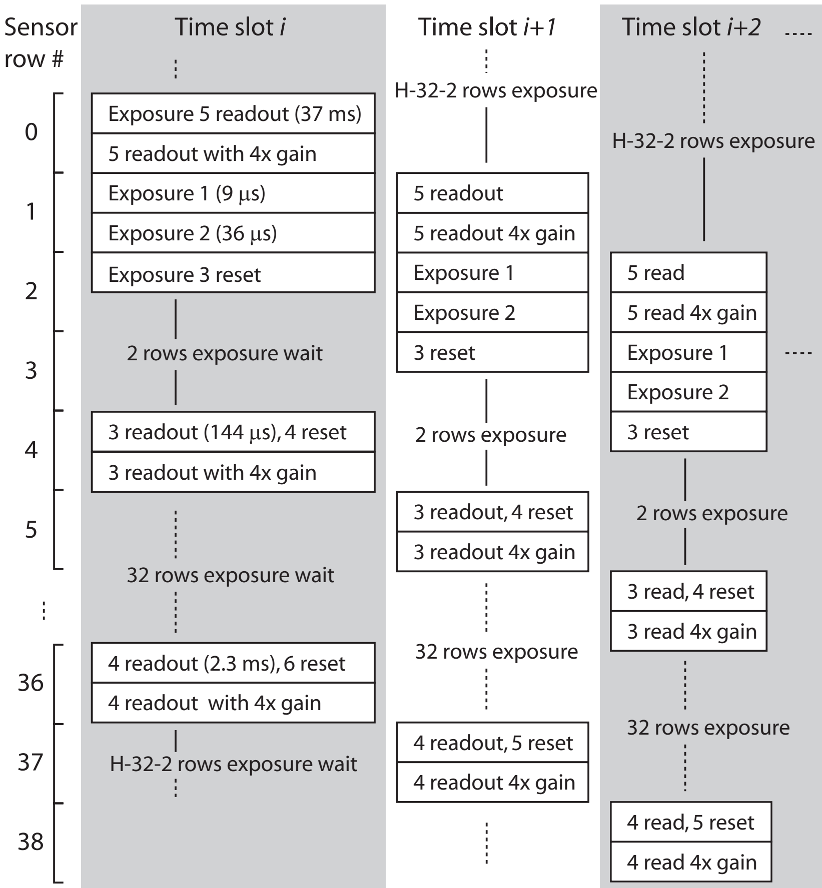


Figure 4: Diagram of the capture algorithm, somewhat simplified for presentation. The progressive image exposure and readout from a rolling shutter algorithm effectively removes any delay between subsequent exposures within each HDR frame. Each time slot is 78 µs, during which resets and readouts of several different sensor rows are performed. One full frame is captured in as many time slots as there are rows in the output image.

Table 1: Exposure parameters for the sample implementation. 8 readouts are performed, but only 5 actual exposures are made.

Image #	Parameters	Image #	Parameters
1	9 µs (exposure 1)	5	2.3 ms (exposure 4)
2	36 µs (exposure 2)	6	4x gain of the above
3	144 µs (exposure 3)	7	37 ms (exposure 5)
4	4x gain of the above	8	4x gain of the above

3.3. Performance

Many factors influence the performance of the system, but the values presented in Table 1 represent a reasonable balance between dynamic range, precision and speed. The useful dynamic range with the parameters shown is around 1,000,000 : 1. An alternative version of our algorithm achieves a 10-fold increase in the dynamic range by using a single-microsecond time for the shortest exposure and placing some of the exposure times wider apart, up to 3 f-stops. For a 25 FPS frame rate and 512 rows of output, 8 exposures are the maximum, but a lower frame rate or fewer rows in the output image would allow for more exposures while keeping the frame rate.

The hard limiting factors for system performance are the sum of all the different exposure times, the internal data transfer rate and data output rate, both around 1 Gbit/s, and the A/D conversion time for each row of pixels. An 8-bit A/D conversion in the ramp converter requires 9.6 µs but, if extra speed is required, a 7-bit conversion can be performed in about half that time. These different factors jointly determine the properties of the imaging system.

Because of the rolling shutter methodology, A/D conversion can be performed simultaneously with exposure. If the number of exposures is N, the exposure time for exposure number i is Δt_i, and the image resolution is H rows of W pixels each, the resulting minimum frame time t_f can be determined by the following equations:

A/D conversion time for one exposure	$t_c = H \cdot 9.6 \cdot 10^{-6}[s]$
Processing time for exposure i	$t_i = max(t_c, \Delta t_i)$
Data transfer time for one exposure	$t_d = H \cdot W \cdot 8 \cdot 10^{-9}[s]$
Total frame time	$t_f = max(\sum_{i=1}^{N} t_i, N \cdot t_d)$

The processing time, t_i, required for exposure number i is the maximum of the A/D conversion time, t_c, for one full frame of H rows and the exposure time, Δt_i, for exposure i. The data transfer time, t_d, is the time it takes to transfer all i exposures over the 1 Gbit/s data link. The frame time, t_f, is the maximum of the processing and data transfer times.

Note that the rolling shutter capture method eliminates a big problem with traditional multiple exposure methods, where the frame time has to be split equally between the exposures and each individual exposure needs to be read out and transferred separately. In our implementation, the data transfer is performed simultaneously for all the exposures, but in an out-of-order fashion. If the system is not limited by data transfer bandwidth, the frame time can in fact be equal to the sum of the exposure times, not N times the longest exposure time.

4. HDR ASSEMBLY

Traditional multiple exposure techniques use averaging to reduce artifacts in the resulting HDR image, so that each HDR pixel is composed of a weighted average of a number of corresponding pixels from several LDR images. Because we have rapid capture with negligible scene motion and a known, calibrated system with direct access to the linear A/D converted values, we have no need for averaging. Instead, for each pixel we simply pick the best value from the available range of exposures, and encode it as a floating-point photometric value. The “best” value is simply the highest unsaturated value, and the HDR pixel value is computed as $E = x_i/\Delta t_i$, where x_i is the A/D converted value from the sensor after shading correction, Δt_i is the corresponding exposure time, and i is the index for the exposure where x_i has its maximum valid value below the saturation level. Other values are either saturated and useless, or they have lower digital values and therefore lower binary precision. Such values would only decrease the accuracy of the data if they were used.

At each readout from the sensor, two A/D converted values for different exposures for the same pixel are available simultaneously. The two shortest exposures are performed on the same row of pixels within the same time slot of 78 µs, and the subsequent 1x and 4x dual gain readouts are also performed in rapid succession. Because at most one of these values will be used for the final HDR image, and because the sensor chip itself has considerable processing capabilities available, we do not transmit every A/D converted value to the PC host. Instead, a simple multiplexing operation is performed on the sensor, so that for each pair of values for one pixel, only the best value is selected for output, and a final 4-bit value is transmitted for each time slot, denoting which of the two exposures from each pair that were selected. By this multiplexing operation, we save some bandwidth compared to the equations presented above and can transmit a higher resolution image (larger W) than what would have been possible otherwise. By implementing this optimisation to its full potential we would only have to transmit 36 bits of the original 64 bits of data for eight exposures. In the current implementation, a total of 48 bits, six bytes, are transmitted for each row of eight 8-bit exposures, thereby saving 25% on bandwidth.

The bandwidth is thereby reduced to a level where all three cameras of the RGB system can be connected to the same host, a standard PC. On the host, a final multiplexing operation is performed in software to select the best value to use for each pixel of HDR output. Once again, the selection is simple: we pick the highest unsaturated value.

With three cameras connected to the same PC, RGB color HDR frames of 640×512 pixels can be streamed to disk with a sustained frame rate of 25 frames per second. The data stream from the three cameras to the PC is around 1 Gbit/s in total, and the data written to disk is around 300 Mbits/s. Both these figures are well within the bandwidth limits of a standard high quality PC.

The continuous streaming to disk does not allow for much processing of the data by the host. Some extra post-processing still needs to be performed off-line, to perform photometric calibration and registration of the images, but that processing is fast, and it could actually be performed in real time if needed by adding a second CPU to the system. Even for the three-camera system where warping is required to put the RGB components in register, real time performance can be achieved if required, by utilising graphics hardware for the resampling operations. We have implemented such real time resampling in a GPU shader for a real time RGB viewfinder window, but for processing of the final data we still use software methods for simplicity and configurability.

5. QUALITY

This system was designed for high speed, high quality light probe capture. For that purpose, it works very well. However, there are some limitations which could be important for other imaging applications. We first present the advantages of the system, then the limitations.

5.1. Benefits

Because the sensor uses standard CMOS photodiodes, a very mature technology, the image quality of each separate LDR exposure is excellent, and the resulting HDR image quality is also excellent.

The monochrome sensor requires external color filters for RGB capture, but on the plus side, this makes it possible to use color filters with any desired spectral properties. The capture is not even restricted to RGB trichromatic capture, hyper-spectral imaging is also possible.

Because the sensor pixels are fairly large (9 µm square), the light sensitivity is good, and even the relatively short maximum exposure time of 37 ms yields good images even in fairly dim surroundings. The sensor is more sensitive to long wavelengths than short, but the sensitivity to blue light is still good enough for high quality image acquisition in normal indoor lighting conditions.

The sensor was designed for laser imaging applications and gracefully handles extreme local variations in image intensity, with very little blooming to neighboring pixels and no bleeding over large distances. Direct views of the naked sun can be adjacent to dark shadow regions in the image without problems.

Having direct access to the linear A/D converted values and very accurate microcode timing makes the system particularly easy to calibrate. Contrary to most methods where regular LDR cameras are used, we do not need to know and compensate for a nonlinear response curve, nor do we have to recover any uncertain estimate of such a curve.

The HDR assembly from 8 exposures 2 f-stops apart is fully comparable to current practice in still image multi-exposure methods. The photometric accuracy is better than 2 % within a very wide dynamic range. The quantization error for the dynamic range in our example implementation is shown in Figure 5.

The contrast range of our HDR captures can be huge, up to 10,000,000 : 1. This is in fact so large that the limiting factor for the attainable dynamic range is now the flare and glare properties of the optics, not the sensor or algorithm as such.

Finally, as was presented in section 3, the system is highly configurable in software, and trade-offs can be made between dynamic range, precision, image resolution and frame rate to suit a variety of different purposes.

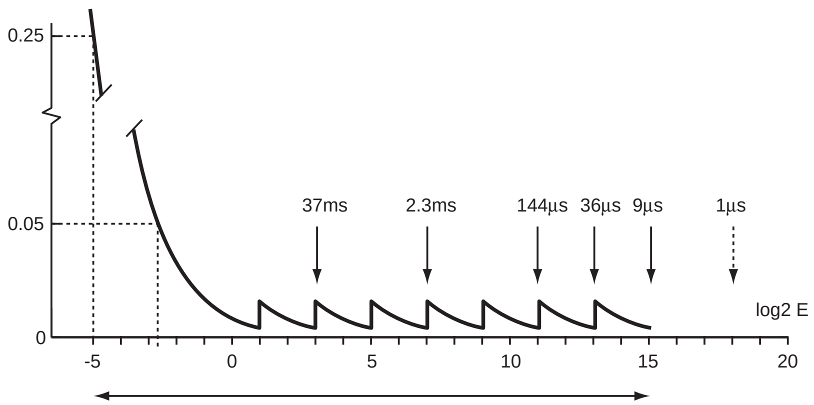


Figure 5: Relative quantization error for the example implementation (black, thick curve). The dynamic range of the demo image on page 1 is indicated at the bottom. The dynamic range of the capture can be extended further by making the shortest exposure 1 µs. More exposures could be added also at the low end of the scale, but that would require a lower frame rate than the current 25 frames per second. Also, while quantization is the dominant source of error for high intensities, low intensities are also affected by thermal noise, which limits the attainable image quality for longer exposures unless the sensor is cooled.

5.2. Limitations

In order to actually obtain a photometrically accurate image with a contrast ratio exceeding a million to one, the lens must have very good internal anti-reflex coatings and good stray light trapping properties. To some extent, flare and glare can be compensated for in software, but high quality optics are required for extreme dynamic range imaging.

Because the capture is performed in a rolling shutter fashion, there is a vertical curtain effect in the image during camera and scene motion. This is a problem for general video applications, but our particular application in image based lighting concerns photometry in a panoramic view, so for us this issue is of no concern.

The large variation in exposure times makes motion blur manifest itself differently in short and long exposures. Combined with the curtain effect, very rapid camera and scene motion in certain unfavourable directions could cause errors in data for some pixels in some frames. Moreover, the multiple exposures are taken at closely spaced but different points in time, which could also be a problem for very rapid scene motion. As suggested in [7], a motion compensation algorithm could be applied to alleviate these effects, but again, our application is not affected by this to any significant amount, as such extremely rapid scene motion does not happen in our panoramic images. For regular video imaging applications, these effects need to be investigated further.

6. CONCLUSION

From our experiences with using the system we have designed, we feel that all our photometry requirements for image based lighting are very well fulfilled. No other current system would give us anywhere near the same speed, control, dynamic range, precision and image quality all at once. The system is a prototype, but it is robust and very easy to use.

For general-purpose HDR video capture, extra software algorithms would be desirable to compensate for the more prominent motion effects. However, it should be noted that even though the curtain effect and the remaining time disparity problems can not be entirely eliminated in our capture, they can be reduced significantly by using a faster capture scan over the sensor. This can be achieved through either a lower output image resolution or through fewer exposures, at the cost of a reduced precision and/or a smaller dynamic range.

We believe that this camera system, and future designs using the same approach, have applications far beyond its current use in our lab, and we would like to point out that HDR video is no longer a dream for the future which requires expensive hardware development projects to be undertaken. It is perfectly possible to do it now, using commercially available camera hardware.

7. APPLICATIONS AND FUTURE WORK

Rapid capture of HDR images was a prerequisite for our plans for high-resolution spatial sampling of illumination information for real world scenes, and that will be the focus of our future efforts. Image based lighting can be improved and generalised significantly by using rapid HDR capture to measure spatially and/or temporally resolved illumination data from real world scenes, such as an object moving through a complicated light field, or a large area with a complex configuration of lights and shadows. The 25 FPS streaming HDR video allows for a more than 100-fold increase in capture speed compared to current practice using still images, and this makes it feasible to sample a light field at densely spaced locations along a path, even over an entire area or within a volume. Such densely sampled light fields provide a very good foundation for highly realistic image based lighting, where the current constraint of spatially uniform lighting conditions is lifted. Each point on a surface can be rendered using different captured points and directions from the light field data set, and variations in illumination conditions across an object can be captured accurately. Our initial work in this area was recently published in [9].

Now that HDR video is possible, there is a clear need for standardised file formats to store, process and share the data. Some efforts are currently being made by others in this field [11], but the dynamic range of currently proposed HDR video formats is aimed at direct display, not measurements of incident illumination. Using HDR images for illumination capture places extra demands on both the range and accuracy of the data. For the time being, we use numbered sequences of uncompressed HDR still images to store video streams.

ACKNOWLEDGEMENTS

We would like to extend our thanks to professor Anders Ynnerman for his active support of this project, and to Mattias Johannesson and Anders Murhed at SICK IVP for fruitful collaboration. This project was supported by the Science Council of Sweden through grant VR2002 621-2001-2623.

REFERENCES

[1] E. Reinhard, G. Ward, S. Pattanaik and P. Debevec, 2006: *High Dynamic Range Imaging - Acquisition, Display and Image-Based Lighting.* Morgan Kaufmann, San Francisco, CA

[2] B. C. Madden, 1993: *Extended intensity range imaging.* Tech. rep., GRASP Laboratory, University of Pennsylvania.

[3] S. Mann and R. W. Picard, 1995: *Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures.* In Proceedings of IS&T 46th annual conference, 422–428.

[4] G. Krawczyk, M. Goesele, and H. P. Seidel, 2005: *Photometric calibration of high dynamic range cameras.* Tech. Rep. Research Report MPI-I-2005-4-005.

[5] S. Nayar and T. Mitsunaga, 2000: *High Dynamic Range Imaging: Spatially Varying Pixel Exposures.* In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, 472–479.

[6] J. Waese and P. Debevec, 2002: *A real-time high dynamic range light probe.* In Proceedings of ACM SIGGRAPH2002, ACM Press/Addison-Wesley Publishing Co., 247.

[7] S. B. Kang, M. Uyttendaele, S. Winder and R. Szeliski, 2003: *High dynamic range video.* ACM Trans. Graph. 22, 3, 319–325.

[8] J. Unger, S. Gustavson, M. Ollila and M. Johannesson, 2004: *A real time light probe.* In Proceedings of the 25th Eurographics Annual Conference, vol. Short Papers and Interactive Demos, 17–21.

[9] J. Unger, S. Gustavson, and A. Ynnerman, 2006: *Densely Sampled Light Probe Sequences for Spatially Variant Image Based Lighting.* In Proceedings of 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia, Kuala Lumpur, Malaysia November 29 - December 02, 2006 (Graphite06), 341 - 347

[10] R. Johansson, L. Lindgren, J. Melander, B. Moller, 2003: *A multi-resolution 100 Gops 4 Gpixels/s programmable cmos image sensor for machine vision.* In Proceedings of the 2003 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, IEEE.

[11] R. Mantiuk, A. Efremov, K. Myszkowski, H. P. Seidel, 2006: *Backward Compatible High-Dynamic-Range MPEG Video Compression.* In Proceedings of ACM SIGGRAPH 2006, ACM Press/Addison-Wesley Publishing Co., 713-723.