A General Approach to Backwards-Compatible Delivery of High Dynamic Range Images and Video

Greg Ward, BrightSide Technologies

Abstract

We propose a general solution to the problem of decoding high dynamic range (HDR) information stored as a supplement to low dynamic range (LDR) images or video. Each LDR frame is paired with a lower resolution HDR version, and these are compressed separately using any of the existing methods appropriate to the task. On decode, the low-resolution HDR image is upsampled to match the resolution of the LDR version, and the high frequencies are transferred from the LDR to the HDR frame. The recovery process places no constraints on the color space or tone-mapping of the backwards-compatible LDR content, and is thus ideally suited to applications such as DVD movies that target legacy equipment while building in forward-compatibility with emerging HDR systems. A fast and simple recovery algorithm is demonstrated, followed by a more sophisticated and accurate technique. Examples are shown on computer-generated video frames as well as HDR captured video.

Background

Recent work in high dynamic range image encoding has focused on "lossy" compression, with a particular emphasis on backwards compatibility with existing formats [Spaulding et al. 2003] [Ward & Simmons 2004, 2005] [Li et al. 2005] [Mantiuk et al. 2006]. This is an important trend for the practical adoption of scene-referred high dynamic range (HDR) digital imagery in a world where lossy, output-referred formats such as JPEG and MPEG dominate. The consumer market will not accept HDR formats that take 12 times as much space as JPEG and 80 times as much space as MPEG, especially if they cannot be displayed with standard viewers and players. Such is the case for the most common HDR formats in use today: *Radiance* RGBE (.hdr), OpenEXR, and TIFF [Reinhard et al. 2005]. Though adoption of such lossless formats is taking place in high end tools such as Adobe Photoshop™ and systems such as Mac OS X and (hopefully) the next version of Windows, the space requirements of lossless HDR will preclude its use in digital photography, video, and web applications for the foreseeable future. Lossy HDR encodings that are not backwards-compatible [Mantiuk et al. 2004] [Xu et al. 2005] will eventually make it to the marketplace, but currently offer no practical transition path.

Methods for backwards-compatible HDR image encoding can be divided into two categories: reversible tone-mapping [Li et al. 2005] and supplemental images [Spaulding et al. 2003] [Ward & Simmons 2004, 2005] [Mantiuk et al. 2006]. Reversible tone-mapping presents a challenge to efficient encoding, because JPEG and MPEG tend to degrade information that is important for proper reconstruction. Thus, compression performance is much worse than with supplemental methods. Supplemental methods encode additional image data to recover the HDR original from the recorded low dynamic range (LDR) information, storing the extra data in an auxiliary stream that is ignored by naïve viewers and players. The key is to minimize the size of this auxiliary stream, and existing methods add between 5% and 30% to the LDR image size, depending on the method and settings.

The problem with supplemental encoding schemes is their inherent complexity. Kodak's ERI format [Spaulding et al. 2003] uses a residual image with sophisticated color and bit manipulations to minimize the size of the auxiliary stream, achieving good compression but only modest gains in dynamic range. The backwards-compatible HDR version of MPEG introduced by Mantiuk et al. [2006] follows a similar approach, with a residual image storing the difference between a predictor function on the LDR data and an perceptual HDR color space. In both methods, the residual tends to be small and noise-like where the LDR image is within its output-referred gamut, but jumps abruptly wherever the LDR image saturates at the top end. This challenges standard lossy image compression techniques, which must be tailored to encode only the perceptually important information without introducing false contours at the gamut boundaries. The different bit sizes between the LDR and HDR data pose additional difficulties during encoding and decoding, and care must be taken not to introduce new quantization errors in the process. To avoid these complications, the JPEG-HDR encoding of Ward & Simmons [2004, 2005] employs a ratio image in place of a residual, which can be multiplied against the decompressed LDR image to recover the HDR original. This simplifies the process by allowing a single 8-bit log channel to store the ratio between HDR and LDR pixels, but complexities creep back in when the ratio image is downsampled to reduce the size of the auxiliary stream. With ratio image downsampling, the LDR image must either be "precorrected" against lost resolution, or "postcorrected" using a resolution enhancement technique. Also, the LDR image must encode all the necessary color information, since the ratio image is only for the luminance channel. This places important restrictions on the tone-mapping operator and color space of the LDR data. Such restrictions are avoided in the method of Mantiuk et al. [2006], which encodes color information as well as HDR luminance in its residual image.

The method we propose for backwards-compatible encoding of HDR imagery is inspired by previous work and motivated by the following observations:

- In today's applications, the LDR data is more important than the HDR data, and should not be compromised.
- HDR imagery takes longer to decompress because hardware and software are tailored to 8-bit streams.
- Both of these conditions will change in the next 5 years.

The main reason that no one has implemented the obvious solution of storing a full-blown HDR image in an auxiliary stream is that it would more than double the data size, incorporating largely redundant information. However, storing a low-resolution version of the HDR original has a number of advantages relative to our three observations. First, it would not affect the LDR data in any way. Second, it would decompress quickly because most algorithms are proportional to the number of pixels – using ¼ the resolution in each dimension would speed decompression by a factor of 16. Third, as more applications come to rely on HDR imagery and the associated hardware and software get faster, the resolution of the auxiliary stream can be increased over time, eventually reversing the roles and importance of the LDR and HDR data streams.

This is the essence of our proposal: along with the conventional LDR image stream (still image or video sequence), we store a corresponding HDR image stream at a reduced resolution. No special preparation is made on either stream, and no restrictions are placed on the color space or tone-mapping of the LDR data. Storing an HDR image with every LDR image may seem redundant, but only the low frequency information is repeated, and from this we can derive a correlation to better recover the high frequencies; so it is not redundant, but necessary. We retain the nicer features of the supplemental encoding methods, with none of the complexity – at least none on the encoding side. Decoding is another matter, and the subject of this paper.

We start by describing a basic recovery method that is fast but depends on a global tone-mapping operator, then describe a more advanced method that estimates the local tone-mapping response automatically. In our results section, we evaluate the performance of our two algorithms on a variety of source images and video, both synthetic and captured, and mapped to different color spaces using different tone-mapping operators. We conclude with some final observations and suggestions for future work.
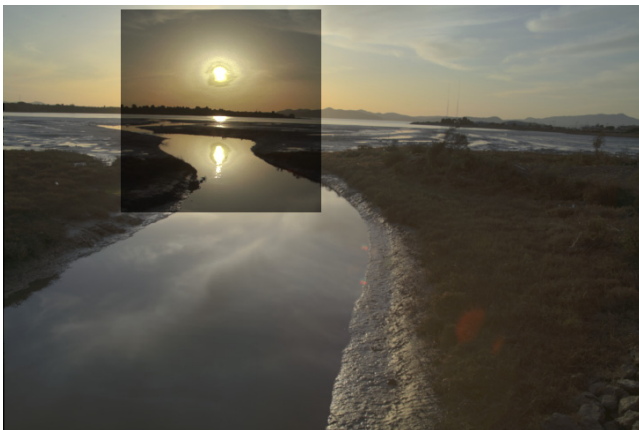

Figure 1. Tone curve inversion from 8-bit/channel original, showing quantization artifacts in bright region.

Method

Our decoder is given two versions of our image, one that is high resolution but low dynamic range, and one that is low resolution but high dynamic range. From these, we wish to derive an image that is high resolution and high dynamic range. If a global tone-mapping operator were used to generate the LDR image, we could try inverting this mapping to arrive at an HDR version, ignoring the low-resolution HDR information provided. However, we would run out of color resolution in places where the image values were clipped (out of gamut) or the tone curve underwent a large degree of compression. Such a result is shown in Figure 1. The sky region surrounding the sun shows quantization artifacts as a result of the expansion of LDR data. Clearly, we need the HDR data to supplement our results where the LDR image is inadequate.

Basic Method

Since we have a low resolution version of our HDR image, we can take a different approach. Rather than inverting the tone curve, we can take the high frequency data above the quantization threshold from the LDR image to augment the low frequency information in the HDR image. Specifically, we:

1. Convert our LDR image color space to approximately match the primaries of the HDR data.
2. Extract high frequencies from the LDR image between HDR and LDR image resolutions and apply quantization threshold.
3. Upsample the HDR image to LDR resolution and apply the high frequencies from Step 2.

The high frequency ratio image can be computed using a rational convolution filter, or with equal efficiency via a downsampling-upsampling-divide process:

A. Downsample the image to lower resolution limit.
B. Upsample again to original resolution.
C. Divide the original image by resampled image from Step B.

This results in a scaling image whose pixels have an average value of 1.0, which can be multiplied by an upsampled image to recover the missing high frequencies. It is critical that the upsampling method not introduce spurious high frequencies, therefore bilinear interpolation is preferred over the more usual bicubic basis in Step B. (We employed a separated Gaussian kernel with a radius of 0.6 pixels in a square footprint of 5x5 pixels for downsampling.) Figure 2 shows a high frequency image before and after a threshold of 1.5 quantization steps is applied.

If we can invert our tone-mapping curve prior to Step 1 above, our results will be fairly accurate because we only take high frequency edge information from the LDR version, avoiding the quantization artifacts that were showing up in the smooth gradient regions. Figure 3 shows the comparison results, recovered from a 480x640 pixel LDR and a 120x160 HDR image.

Unfortunately, we do not always know the tone-mapping that was applied to arrive at the LDR image, and even when we do, it may be too difficult or too expensive to invert, leaving us with inaccuracies in our high frequency data. Globally, our recovered

images will look about right, but details may be softened or lost, as shown in Figure 4. In general, tone-mapping operators may be globally or locally determined, and may preserve visibility over the entire image or may allow some regions to saturate to black or to white. In saturated regions, we must fall back on the low frequency information in our HDR image. This is acceptable in most cases, as this information was selected out of the LDR version by the content creators.
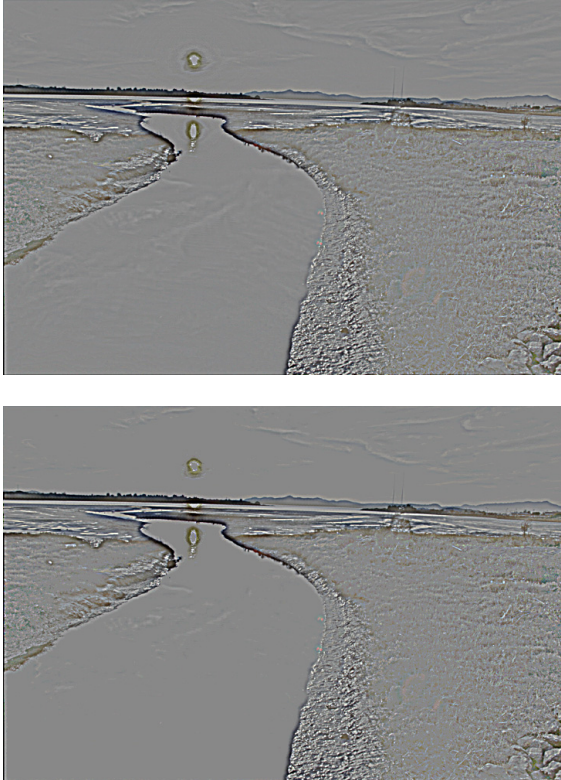

Figure 2. High frequency image before and after quantization threshold is applied.

Advanced Method

The basic method just described is very similar to the "postcorrection" technique introduced by Ward & Simmons [2004], which is known to be a crude approximation due to its ignorance of the tone-mapping function. When the LDR tone-mapping is complex or unknown, we still wish to make the best use of high frequencies possible. To accomplish this, we need to approximate the impulse response in the LDR data, which is the important part of the tone-mapping for our purposes. Because the tone-mapping operator can vary over the image, we need an approximation of the impulse function that is also allowed to vary. But how do we derive such an approximation? We need to relate our LDR and HDR data, but we cannot compare the frequencies we wish to recover, since they are missing from our HDR image. So, we settle for the closest information available – the top frequencies in the lower resolution image, assuming the impulse response does not change dramatically from one band to the next.[1]

---

[1] This assumption may be violated in tone-mapping operators that incorporate a sharpening stage. In such cases, we can either undo

Figure 3. Basic method of recovering high frequencies using a priori knowledge of the tone-mapping function.

Figure 4. Basic method without knowledge of tone-mapping curve, showing loss of detail relative to Figure 3.

The steps are listed below:

1. Convert our LDR image to match the color space of the HDR image, I_{SR}. Call this I_{SOR}.
2. Reduce to the resolution of I_{SOR} to match that of I_{SR}. Call this I'_{SOR}.
3. Extract high frequencies for grayscale versions of I_{SR} and I'_{SOR} using a bandwidth equal to the resolution difference between I_{SR} and I_{SOR}.
4. Compute a spatially varying impulse response function between the I'_{SOR} upper band image and the I_{SR} upper band image. (Estimating the impulse response is explained in the following section.)
5. Compute the highest frequencies of full-resolution image I_{SOR} and apply the impulse response from Step 4 to recover the HDR high frequencies.
6. Upsample I_{SR} to match the full LDR image resolution.
7. Multiply high frequency data from Step 5 to get our full-resolution HDR output.

---

the sharpening filter prior to recovery or accept that our final HDR output will exhibit a similarly sharpened result.

Applying our more advanced method to the same tone-mapped image as before, we see improved sharpness in the unsaturated regions of Figure 5, based on the same input used to generate Figure 4. The additional high frequency and mapping calculations increase the processing time by a factor of 4, from 0.45 to 1.8 seconds on a 1.5 GHz G4 processor, with both calculations proportional to the number of pixels.

Figure 5. Advanced method estimates the high frequency impulse response over the image, obtaining a sharp result without assumptions about the tone-mapping operator.

Figure 6 shows a comparison between the simple method and the advanced method on a 2704x4064 LDR image mapped using Durand & Dorsey's bilateral filter [2002], paired with a 676x1016 HDR version. The blue inset shows the red area recovered using the simple method without knowledge of the TMO. The green inset shows the same area mapped using the advanced method to estimate the TMO. Although neither reconstruction is perfect, the advanced method retains greater sharpness, at the expense of some over-shooting, visible as slight discolorations at high contrast edges. For comparison, the white inset shows the HDR data upsampled with a bicubic filter on the left, and the full-resolution original on the right. (Insets were tone-mapped using a histogram operator [Larson et al. 1997].)

Estimating the Impulse Response Function

In order to correctly map the high frequencies in the tone-mapped LDR image into the HDR domain, we need to estimate the local impulse response. We start by assuming that the impulse response function is monotonically increasing, at least locally. This is a reasonable assumption for any tone-mapping operator, since a decreasing impulse response would imply that larger gradients in the original yield smaller gradients in the tone-mapped image, which would show up as edges with reversed contrast. Even with this assumption, it is difficult to estimate a continuously changing function, so we further assume that the function is constant within small local regions of the image. In our implementation, we use overlapping blocks of roughly 64x64 pixels in the subsampled HDR input, regardless of the input image resolution. We found this to be a reasonable size to obtain a sampling of the impulse response.

Figure 6. An image of the Stanford Memorial Church mapped with a bilateral filter, and recovered using the simple and advanced methods from a 4x downsampled HDR version. (Image courtesy Chris Cox of Adobe Systems.)

Within each block, we put the luminance (gray) values from the LDR upper band image in one 4096-entry data array corresponding to 64x64 pixels, and the luminance values from the HDR upper band image in a second data array. We independently sort the two arrays, then extract the input-output pairs corresponding to every 164th entry. By construction, this creates a set of 25 monotonically increasing coordinates, evenly spaced in the data population, which we can use in a linear or cubic interpolation of the impulse response function. The graph in Figure 7 show the original scatter of values from an example HDR/LDR mid-frequency block pair in Figure 6, and the sorted interpolation points. In regions such as the one chosen here, we may get diverging impulse responses that we combine into a single curve, but this is necessary in order to derive a smooth function of one variable.[2] Besides enforcing monotonicity, independent sorting avoids outliers caused by minor misalignments between the LDR and HDR data. Computing a single impulse response function based on luminance further minimizes color shifts when we apply it independently to each channel during reconstruction.

---

[2] Better correlations could be obtained by adding a second variable to our scheme, the original HDR luminance. This adds a level of complexity, but could improve the results for some tone-mapping operators.
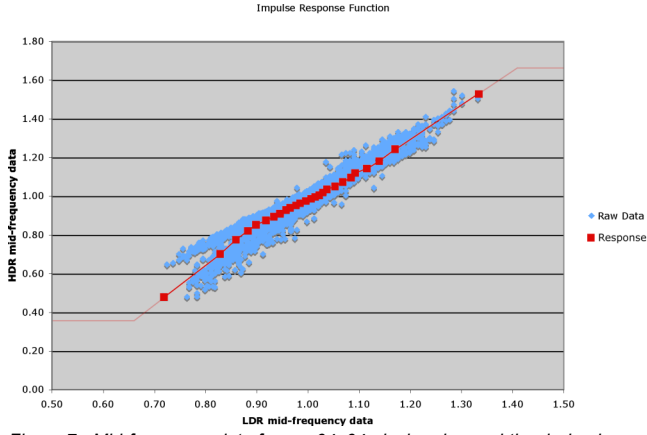
Figure 7. Mid-frequency data from a 64x64 pixel region and the derived impulse response function to be applied to the highest frequencies.

The blocks we use to derive the local impulse response functions overlap, enabling us to smoothly interpolate the results over the image. This scheme is shown diagrammatically in Figure 8. Consider the four neighboring 64x64 blocks: A, B, C, & D. Each has its center at the intersection of four 32x32 pixel regions, and characterizes the response in the surrounding square using the method just described.[3] To map the response of the impulse for pixel **P**, we pass it through each of the four response functions, and linearly interpolate the results based on the position of **P**.

To avoid excessive quantization noise in high gradient areas, we place a restriction on the maximum overall gain for the impulse response within each block. If the difference between the maximum and minimum output values is greater than 5 times the difference between the maximum and minimum input values, the response is scaled to reduce the average slope to fit the maximum 1:5 ratio. This limit is rarely reached in practice.
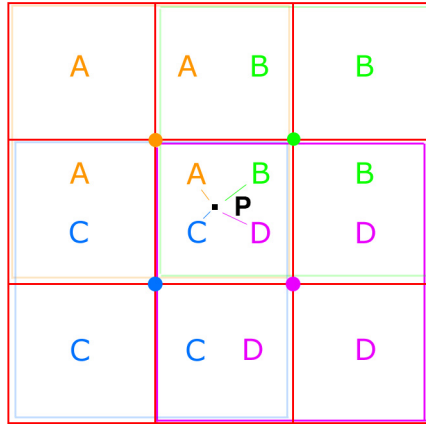

Figure 8. Overlapping impulse response blocks permit smooth interpolation of output.

---

[3] The 50% overlap we have chosen with each neighbor is adjustable. We recommend at least 25% neighbor overlap for smooth results.

It is also important to consider values outside the range of the interpolated response pairs. We extrapolate the last value pair on top and bottom out to one half the distance between the last two points. After that, we cap the impulse response function, effectively cutting off high frequencies this far above the observed gradient values. This response extension is shown as the faded ends on the curve in Figure 7. Limiting extrapolation in this way reduces ringing artifacts caused by sharp edges in the LDR image.

Results

We tested three variants of our algorithm each on a computer-generated HDR animation and two captured HDR video sequences. The first variant of our algorithm assumes a linear mapping between HDR and LDR (tone-mapped) versions. The second variant assumes a non-linear, global tone-mapping over the image. The third variant is our "advanced" method, which allows the tone-mapping to change locally and non-linearly. Figure 9 shows stills from our lighting simulation of an air traffic control tower, a beach at sunset, and a trip through a tunnel.

Figure 9. Our three test scenes.: an HD-resolution animatin and two VGA-resolution video sequences.

Not surprisingly, we found that the linear variant worked well enough on the linear tone-mapping operator, but there was considerable detail lost in the brighter regions where the linear operator had clipped (Figure 10). The linear assumption was not appropriate for any of the other tone-mapping operators, and tended to produce excessive sharpness in the highlights due to over-shooting (Figure 11). Also not a surprise, we found that the global, non-linear algorithm variant was acceptable for the global versions of the Reinhard [Reinhard et al. 2002] and histogram [Larson et al. 1997] operators, but degenerated with the bilateral filter [Durand & Dorsey 2002] and the gradient domain operator [Fattal et al. 2002], due to their local behavior.

On the synthetic control tower animation, the advanced algorithm performed acceptably for every tone-mapping operator we tried, though clamping in the linear operator still lost high frequency

information. In the HDR captured sequences, we noticed a few problems at the boundaries of bright objects, which appeared as colorful outlines. Even within some low-gradient fields, such as the orange sky of the sunset, patterns would occasionally emerge (Figure 12). These are due to the Bayer pattern of the color image sensor, and the best solution is to improve the demosaicing filter. Barring that, a comb filter could be applied during HDR recovery to reduce the appearance of Bayer mosaic remnants.

Conclusion

We have proposed a solution for backwards-compatible HDR imagery that stores a low-resolution HDR version of each frame as a supplement to the LDR data. This places the burden on the decompression engine to recover high-resolution HDR frames by combining the two streams. This approach is not necessarily better or faster than previous backwards-compatible solutions. The real benefit to our approach is the evolutionary path it offers, especially for video.

Any backwards-compatible format is a stop-gap solution that requires compromises in encoding efficiency. This is illustrated by the lower performance of the backwards-compatible HDR extension to MPEG by Mantiuk et al. [2006] relative to their original proposal, which was not backwards-compatible [Mantiuk et al. 2004]. Unfortunately, backwards-compatibility is considered essential to market adoption, especially for video. Furthermore, once we settle on a new encoding standard, we are committed to it for about 10 years. This seems like a long time to be using an ornate, stop-gap format, and in the end we might prefer a simpler solution that provides a smooth transition to native HDR video.

This is exactly what we offer. By logically separating the HDR and LDR data streams, tying the quality of each to its resolution alone, we provide a seamless upgrade path from the LDR world of today to the HDR world of tomorrow. We are free to standardize on the most efficient HDR encoding we can devise, with no compromises for backwards-compatibility. We can then incorporate this standard in new hardware and software, coupling it with a method to extract resolution from the legacy LDR stream. As time goes on, hardware and software will continue to improve, enabling real-time decoding of higher resolution HDR streams, simultaneously obviating the need for LDR data. Eventually, the LDR stream will become subservient to HDR, offering little more than a low-resolution tone-mapping suggestion for legacy devices. Color management will then move into display devices, and a high dynamic range profile connection space will become the preferred delivery medium.

While the basic recovery method described can be implemented efficiently on the GPU and works well enough for global tone-mapping operators, the advanced method relies on accumulation and sorting operations that are more conveniently carried out on the CPU, and is currently too slow for real-time playback. It should be possible to implement the advanced method in a more GPU-friendly way, by replacing our derivation of the impulse response function or off-loading this efficiently. This is left as future work, along with the reverse problem of recovering high frequencies in an LDR image given a high resolution HDR frame. We expect the solution to look very similar to the current one, with better results thanks to the greater bit depth of the HDR stream.

References

[Durand & Dorsey 2002]
> FREDO DURAND & JULIE DORSEY, "Fast Bilateral Filtering for the Display of High-Dynamic Range Images," *ACM Transactions on Graphics*, 21, 3.

[Fattal et al. 2002]
> RAANAN FATTAL, D. LISCHINSKI, M. WERMAN, "Gradient Domain High Dynamic Range Compression," *ACM Transactions on Graphics*, 21, 3.

[Larson et al. 1997]
> GREG WARD LARSON, HOLLY RUSHMEIER, CHRISTINE PIATKO, "A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes," *IEEE Trans. on Visualization and Computer Graphics*, 3, 4.

[Li et al. 2005]
> YUANZHEN LI, LAVANYA SHARAN, EDWARD ADELSON, "Compressing and Companding High Dynamic Range Images with Subband Architectures," *Proc. of SIGGRAPH '05 (Special issue of ACM Transactions on Graphics)*.

[Mantiuk et al. 2004]
> RAFAŁ MANTIUK, GRZEGORZ KRAWCZYK, KAROL MYSZKOWSKI, HANS-PETER SEIDEL, "Perception-motivated High Dynamic Range Video Encoding," *Proc. of SIGGRAPH '04 (Special issue of ACM Transactions on Graphics)*.

[Mantiuk et al. 2006]
> RAFAL MANTIUK, ALEXANDER EFREMOV, KAROL MYSZKOWSKI, HANS-PETER SEIDEL, "Backward Compatible High Dynamic Range MPEG Video Compression," *Proc. of SIGGRAPH '06 (Special issue of ACM Transactions on Graphics)*.

[Reinhard et al. 2002]
> ERIK REINHARD, M. STARK, P. SHIRLEY, J. FERWERDA, "Photographic Tone Reproduction for Digital Images," *ACM Transactions on Graphics*, 21,3.

[Reinhard et al. 2005]
> ERIK REINHARD, GREG WARD, SUMANTA PATTANAIK, PAUL DEBEVEC, *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting*, Morgan Kaufmann Publishers, San Francisco.

[Spaulding et al. 2003]
> KEVIN SPAULDING, G. J. WOOLFE AND R. L. JOSHI, "Extending the Color Gamut and Dynamic Range of an sRGB Image using a Residual Image," *Color Res. Appl.* 28.

[Ward & Simmons 2004]
> GREG WARD & MARYANN SIMMONS, "Subband Encoding of High Dynamic Range Imagery," *First Symposium on Applied Perception in Graphics and Visualization* (APGV).

[Ward & Simmons 2005]
> GREG WARD & MARYANN SIMMONS, "JPEG-HDR: A Backwards-Compatible, High Dynamic Range Extension to JPEG," *Proceedings of the Thirteenth Color Imaging Conference*.

[Xu et al. 2005]
> RULFENG XU, SUMANTA PATTANAIK, CHARLES HUGHES, "High-dynamic range still-image encoding in JPEG 2000," *IEEE Comp. Graph. and Appl.* 26, 6.
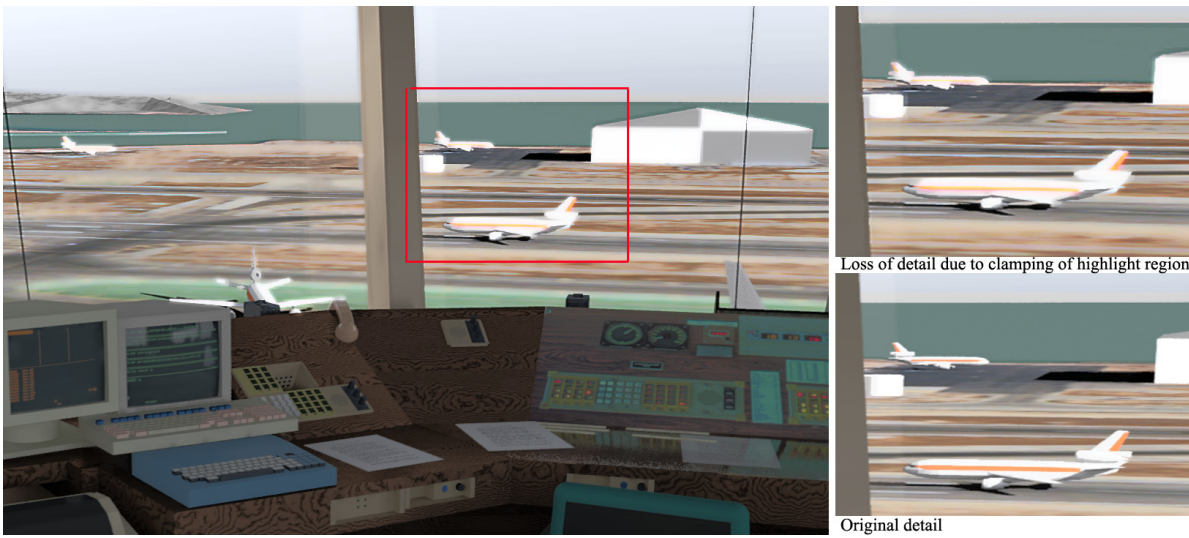
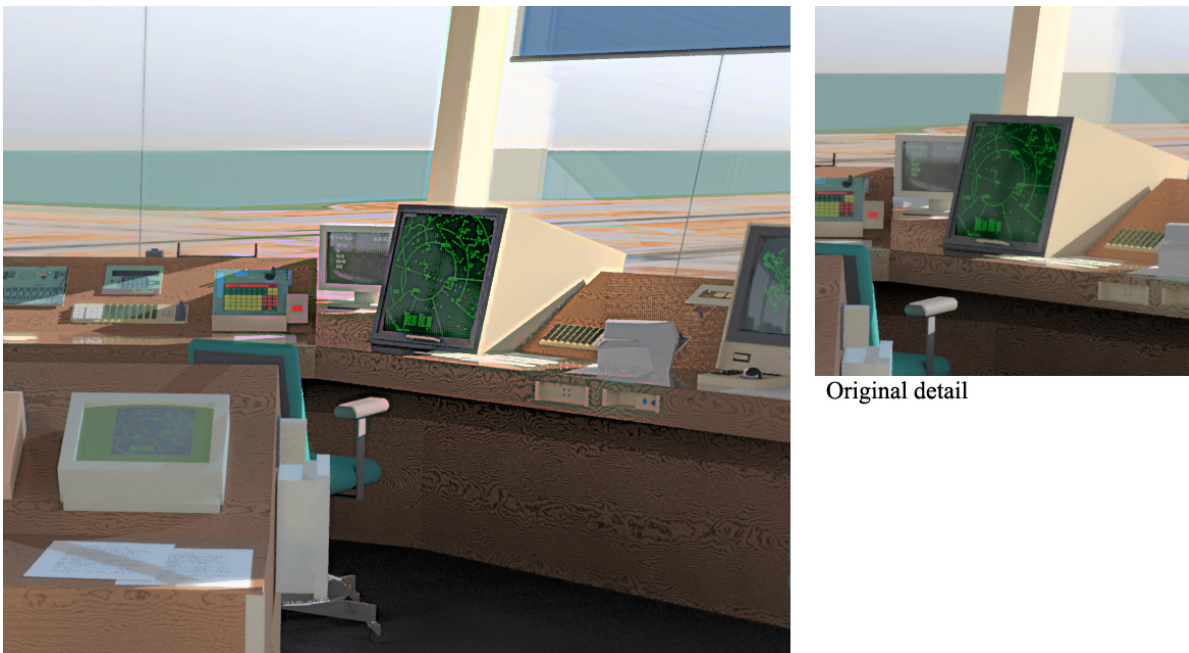Figure 10. Linear tone-mapping clamps highlights, leaving only low-resolution HDR data for reconstruction.

Figure 11. Reinhard tone-mapping of LDR image results in exaggerated high frequencies in some regions if we assume a linear response.

Figue 12. Advanced HDR recovery method can aggravate Bayer mosaic artifacts, as shown in this sunset capture. (Tone-mapped input on right.)