High Dynamic Range Imaging and Low Dynamic Range Expansion for Generating HDR Content

Francesco Banterle 1 Kurt Debattista 1 Alessandro Artusi 1 Sumanta Pattanaik 2 Karol Myszkowski 3
Patrick Ledda 1 Marina Bloj 4 and Alan Chalmers 1

1 The Digital Laboratory, WMG, University of Warwick
2 University of Central Florida
3 Max-Planck-Institut für Informatik
4 Optometry Department, University of Bradford

Abstract
In the last few years, researchers in the field of High Dynamic Range (HDR) Imaging have focused on providing tools for expanding Low Dynamic Range (LDR) content for the generation of HDR images due to the growing popularity of HDR in applications, such as photography and rendering via Image-Based Lighting, and the imminent arrival of HDR displays to the consumer market. LDR content expansion is required due to the lack of fast and reliable consumer level HDR capture for still images and videos. Furthermore, LDR content expansion, will allow the re-use of legacy LDR stills, videos and LDR applications created, over the last century and more, to be widely available. The use of certain LDR expansion methods, those that are based on the inversion of tone mapping operators, has made it possible to create novel compression algorithms that tackle the problem of the size of HDR content storage, which remains one of the major obstacles to be overcome for the adoption of HDR. These methods are used in conjunction with traditional LDR compression methods and can evolve accordingly. The goal of this report is to provide a comprehensive overview on HDR Imaging, and an in depth review on these emerging topics.

1. Introduction
High Dynamic Range (HDR) Imaging has become one of the main areas of computer graphics. One major aspect of HDR imaging which is bound to become extremely relevant is the aspect of providing content for HDR displays. While content can be captured directly for HDR displays [DM97], this is typically not a straightforward process and may require specialised equipment to automate [Sph02, Pan02], just to obtain still images. The provision of animated HDR content is still in its infancy and few reliable methods exist to directly capture HDR video [Hoe07]. This has led to research into providing HDR content from Low Dynamic Range (LDR) originals. Such work makes it now possible to re-use the large amount of already existing legacy LDR in a way that makes full use of emerging HDR displays. Furthermore, several methods, based on LDR to HDR expansion, have been used for HDR compression and for enhancing the quality of rendered images based on HDR image-based lighting (IBL). While previous techniques dealing with general HDR methods have been collected and published, for example [RWP05], at this time only a short survey without in-depth discussion and analysis of algorithmic aspects of LDR expansion techniques has been published by Myszkowski et al. [MMK08]. Myszkowski et al.’s work does not cover the association between LDR to HDR expansion and HDR compression techniques as is presented in this survey.

We begin this state-of-the-art report by giving a quick overview of the different aspects of HDR imaging. In Section 3 we present methods that expand LDR into HDR content with respect to still images, videos and the use of expansion in applications such as IBL. We classify these methods and present the work that has been done to validate such techniques. In Section 4 we show how LDR to HDR expansion methods have been used to compress HDR content, by taking advantage of already existing LDR compression
schemes. Finally, we conclude by discussing open problems and present future directions.

2. High Dynamic Range Imaging

The introduction of HDR imaging in the last two decades by the computer graphics community has revolutionised the field and other areas such as photography, virtual reality, visual effects, and the video-games industry. Physically-correct light values can now be captured and fully utilised for various applications without the need to linearise the signal and to deal with clamped values. The very dark and bright areas of a scene can be recorded at the same time into an image or a video, avoiding under-exposed and over-exposed areas. Traditional imaging methods do not use physical values and typically are constrained by limitations in technology that could only handle 8-bit per colour channel per pixel. Such imagery (8-bit or less per colour channel) is referred as LDR imagery. This change in how light can be recorded is comparable to the introduction of colour photography and has changed each stage of the imaging pipeline, see Figure 1. The four main stages are: capturing, storing, processing, and displaying.

2.1. Capturing

Currently, available consumer cameras are limited to capture only 8-bit images or 12-bit images in RAW format, which do not cover the full dynamic range of irradiance values in most environments in real world. The only possibility is to take a number of exposures of the same scene to capture details from very dark regions to very bright regions as proposed by Mann and Picard [MP95]. The problem with film and digital cameras is that they do not have a linear response, but a more general function \( h \), called camera response function (CRF). Mann and Picard [MP95] proposed a simple method for calculating \( h \), which consists of fitting the values of pixels at different exposures to a fixed CRF, \( h(x) = ax^2 + b \). This parametric \( h \) is very limited and does not support most real CRFs. Debevec and Malik [DM97] proposed a simple method for recovering a CRF through a tabled \( h \) which is minimised using a squared error function. Mitsunaga and Nayar [MN99] improved this algorithm with a more robust method based on a polynomial representation of \( h \). Note that the multiple exposure method assumes that images are perfectly aligned, there are no moving objects, and CCD noise is not a problem. Robertson et al. [RBS99, RBS03] improved previous techniques for assembling HDR images from multiple exposures. They proposed an iterative calculation of the CRF in conjunction with a probabilistic weighting for merging different exposures.

Kang et al. [KUWS03] extended multiple exposure images methods for videos. They presented a system that had a programmed video-camera that temporally varies the shutter speed at each frame. The final video is generated aligning and warping corresponding frames at different shutter speeds and compositing them to recover the HDR one. However, the frame rate of this method is low, around 15 fps, and the scene has to contain low speed moving objects otherwise artifacts will appear. Nayar and Branzoi [NB03] developed an adaptive dynamic range camera, where a controllable liquid crystal light modulator is placed in front of the camera. This modulator adapts the exposure of each pixel on the image detector allowing to capture scenes with a very large dynamic range.

In the commercial field, few companies provide HDR cameras based on automatic multiple exposure capturing. The two main cameras are Spheron HDR VR camera [Sph02] by SpheronVR GmbH and Panoscan MK-3 [Pan02] by Panoscan Ltd, which are both full 360 degree panoramic cameras at high resolution. The two cameras capture full HDR images. For example, Spheron HDR VR can capture 26 f-stops of dynamic range at 50 Megapixels resolution in 24 minutes.

The alternative to automatic multiple exposure cameras is to use CCD sensors which can natively capture HDR values. In recent years, CCDs that record into 10/12-bit per channel in the logarithmic domain have been introduced by many companies such as Cypress Semiconductor [Cyp], Omron [Omr07], PTGrey [PtG04], Neuricam [Neu06], etc. The main problem with these sensors is that their resolution is low, VGA (640 × 480), and are noisy. Therefore, their applications are oriented to automotive, security, and automation use in factories.

In the cinema industry a few companies have proposed high quality solutions such as Viper camera [Tho05] by Thomson GV, Red One camera [Red06] by RED Company, and the Phantom HD camera [Vis05] by Vision Research, etc. All these video-cameras present high frame rates, low noise, high definition (1,920 × 1,080) resolution, and a good dynamic range (reaching the range of celluloid film), 10/12-bit per channel in the logarithmic/linear domain. However, they are extremely expensive (sometimes available only for renting) and they do not encompass the full dynamic range of the Human Visual System (HVS).

2.2. Storing

Once HDR images/videos are captured from the real world, or are synthesised using computer graphics, there is the need to store, distribute, and process these images. An uncompressed HDR pixel is represented using three single precision floating point numbers [Hou01], assuming three bands as for RGB colours. This means that a pixel uses 96 bits per pixel (bpp). Researchers have been working on compression methods to address the high demand on memory storage required for HDR content.

The early compression methods proposed an efficient and compact representation of floating point numbers, the main...
formats are: RGBE/XYZE, LogLuv, and half precision numbers. RGBE/XYZE [War91] is an implementation of floating point where the exponent of the floating point is shared between RGB or XYZ values assuming the exponents have a similar magnitude. For this method the storage requirement is 32 bpp. LogLuv method [Lar98] proposed to separate luminance and chrominance where luminance is encoded in the logarithmic domain achieving 24/32 bpp. Finally, the OpenEXR file format [Ind02] proposed the use of half precision numbers, a 16-bit version of IEEE-754 standard [Hou81], maintaining the dynamic range in an extremely high quality.

LDR Imaging compression methods have also been extended to HDR. For example, block truncation methods in separate luminance and chrominance colour spaces [MCHAM06, RAI06, SLWL08] were applied to HDR textures, achieving 8 bpp. Moreover, Wang et al. [WWS’07] proposed a separated encoding for the HDR and LDR parts of a texture at 16 bpp. Other methods that exploit tone mapping, inverse tone mapping, and LDR de facto standards. They are reviewed in Section 4.

2.3. Image-Based Lighting

HDR content can simplify the definition and rendering process for re-lighting synthesised objects. In particular, IBL techniques are aimed at simulating light transport, defining light sources and surrounding environment. Blinn and Newell [BN76] first used IBL to simulate the optical effects such as reflection and refraction. This was extended by Miller and Hoffman [MH84] and Green [Gre86] for simulating diffuse effects by convolving environment maps. However, these methods were limited to pure diffuse or pure specular materials, without taking into account visibility or secondary rays.Debevec [Deb98] generalised these techniques using ray tracing framework applying the Rendering Equation [Kaj86]:

\[
L(x, \omega) = L_e + \int_{\Omega} L(\omega') f_r(\omega', \omega) V(x, \omega') n \cdot \omega' d\omega'
\]

where \(x\) and \(n\) are respectively the position and normal of the hit object, \(L_e\) is the emitted radiance at point \(x\), \(L\) is the environment map, \(f_r\) is the BRDF, \(\omega'\) is the out-going direction, and \(\omega\) is the view vector. \(V\) is the visibility function, a Boolean function that determines if a ray is obstructed by an object or not. This technique can be applied to real-world objects or human beings for re-lighting them using HDR content [DHT’00]. Therefore, re-lighting using HDR images/videos is a very important application in many fields such as augmented reality, visual effects, and computer graphics. This is because the appearance of the image is transferred onto the re-lighted objects.

\[\text{Figure 1: The HDR pipeline in all its stages. Multiple exposure images are captured and combined obtaining an HDR image. Then this image is quantised, compressed, and stored on the hard disk. Further processing can be applied to the image. For example, areas of high luminance can be extracted and used to re-light a synthetic object. Finally, the HDR image or processed ones can be visualised using traditional LDR display technologies or native HDR monitors.}\]
2.4. Displaying

Most of the display devices commercially available nowadays are not able to display HDR content. This is because current display technology has a low contrast ratio around 1,000 : 1 and can process only 8/10-bit images for each colour channel. In the last two decades researchers spent significant efforts to compress the range of HDR images and videos in order to display them on LCD displays. Tone mapping is the operation that reduces the dynamic range of the input content to fit the dynamic range of the display technology. One of the important requirement is that this reduction of the range must preserve some properties of the original content such as local and global contrast, opponent brightness, etc. Tone mapping is performed using an operator, $f$, which is referred to as a Tone Mapping Operator (TMO). TMOs can be classified in different groups based on the underlying image processing techniques. The main groups of the taxonomy are: Global Operators (the same mapping, $f$, is applied to all pixels), Local Operators (the mapping, $f$ of a pixel depends on its neighbour pixels), Segmentation Operators (the image is segmented in broad regions, and a different mapping is applied to each region), Frequency/Gradient Operators (low and high frequencies of the images are separated, while an operator is applied to the low frequencies, high frequencies are usually retained to preserve fine details), Perceptual Operators ($f$ models some aspects of the HVS), Empirical Operators ($f$ tries to create pleasant images inspired by other fields such as photography), and Temporal Operators (suitable for tone mapping HDR videos). For an in-depth review on tone mapping see Reinhard et al. [RWPD05].

Only in the last few years, researches have been working on display technologies for a native visualisation of HDR images and videos without using TMOs. The two main devices are the HDR viewer [SHS04] and the HDR Monitor [SHS04]. Internally, both of these devices divide an HDR image into a detail layer with colours and a luminance layer that back-modulates the first one.

3. LDR to HDR Expansion

The capture of HDR via multiple exposures using a traditional camera is a very time consuming task, and on a movie set the time for capturing images is very limited. More-
3.2. Global Models

Global models are those methods that apply the same single global expansion function on the LDR content at each pixel in the entire image.

3.2.1. A Power Function Model for Range Expansion

One of the first expansion methods was proposed by Landis [Lan02]. This expansion method, used primarily for IBL, is based on power functions. The luminance expansion is defined as:

\[
L_w(x) = \begin{cases} 
(1-k)L_d(x) + kL_w,\text{Max}L_d(x) & \text{if } L_d(x) \geq R \\
L_d(x) & \text{otherwise}
\end{cases}
\]

\[k = \left( \frac{L_d(x) - R}{1-R} \right)^\alpha\]

where \(R\) is the threshold for expansion which is equal to 0.5 in the original work, \(L_w,\text{Max}\) is the maximum luminance which the user needs for the expanded image, and \(\alpha\) is the exponent of fall-off that controls the stretching curve.

While this technique produces suitable HDR light-probes.
for IBL, see Figure 4, it may not produce good quality images/videos that can be visualised on HDR monitors. This is due to the fact that it does not handle artifacts such as exaggeration of compression or quantisation artifacts.

3.2.2. Linear Scaling for HDR Monitors

In order to investigate how well LDR content is supported by HDR displays, Akyüz et al. [AFR07], run a series of psychophysical experiments. The experiments were run to evaluate tone mapped images, single exposure images and HDR images using the DR-37P HDR monitor. The experiment involved 22 naïve participants between 20 and 40 years old, and in all experiments 10 HDR images ranging from outdoor to indoor, from dim to very bright light conditions were used. The HDR images had around 5 orders of magnitude in order to be mapped to the DR-37P Dolby HDR monitor [Dol05].

The first experiment was a comparison between HDR and LDR images produced using various TMOs [LRP97, DD02, RSSF02], an automatic exposure (that minimises the number of over/under-exposed pixels), and an exposure chosen by subjects in a pilot study. Images were displayed on the DR-37P, using calibrated HDR images and LDR images calibrated to match the appearance on a Dell UltraSharp 2007FP 20.1" LCD monitor. Subjects had the task of ranking images which were looking best to them. For each original test image a subject had to watch a trial image for 2 seconds which was randomly picked between the different type of images. The experimental results showed that participants preferred HDR images. The authors did not find a large difference in participant preference between tone mapped and single exposure (automatic and chosen by the pilot) images.

In the second experiment the authors compared expanded single exposure with HDR and single exposure images (automatic and chosen by the pilot). To expand the single exposure images, they employed the following expansion method:

\[
L_w(x) = k \left( \frac{L_d(x) - L_{d, \text{Min}}}{L_{d, \text{Max}} - L_{d, \text{Min}}} \right) ^{\gamma}
\]

where \(k\) is the maximum luminance intensity of the HDR display, and \(\gamma\) is the non-linear scaling factor. For this experiment images with different \(\gamma\) values equal to 1, 2.2 and 0.45 were generated. The setup and the ranking task was the same as the first experiment. The results showed that brighter chosen exposure expanded images were preferred to HDR images, and vice versa when they had the same mean luminance. Authors suggested that mean luminance is preferable to contrast. Finally, another important result is that linear scaling, \(\gamma = 1\), was the most favoured expansion, suggesting that a linear scaling may be enough for an HDR experience.

The authors worked only with high resolution HDR images, without compression artifacts, and artistically captured. While this works well under such ideal conditions, in more realistic scenarios, such as television programmes or DVDs, where compression is employed, this may not always be the case. In these cases a more accurate expansion needs to be done in order to avoid amplification of compression artifacts, and contouring.

3.3. Classification Models

The methods of Meylan et al. [MDS06, MDS07] and Didyk et al. [DMHS08] attempt to expand different aspects of the LDR content by identifying or classifying different parts in the image such as highlights and light sources.

3.3.1. Highlight Generation for HDR Monitors

Meylan et al. [MDS06, MDS07] presented an inverse Tone Mapping Operator (iTMO) with the specific task of representing highlights in LDR images when displayed on HDR monitors. The main idea is to detect the diffuse and specular part of the image and to expand these using different linear functions. The detection is based on the assumption that highlights are small and bright, which means that the maximum diffuse luminance value \(\omega\) is obtained as the maximum of the low-pass filtered luminance channel \(L_d\) of the image. However, more processing is needed to avoid the case when white diffuse regions are next to regions with highlights, see Figure 5 for the complete pipeline for calculating \(\omega\).

After the calculation of \(\omega\), the luminance channel is expanded using the following function:

\[
f(L_w(x)) = \begin{cases} 
    s_1 L_d(x) & \text{if } L_d(x) \leq \omega \\
    s_1 \omega + s_2 (L_d(x) - \omega) & \text{otherwise}
\end{cases}
\]

\[
s_1 = \frac{\rho}{\omega}, \quad s_2 = \frac{1 - \rho}{L_{d, \text{Max}} - \omega}
\] (1)

where \(L_{d, \text{Max}} = 1\) since the image is normalised, and \(\rho\) is the percentage of the HDR display luminance allocated to the diffuse part. However, a global application of \(f\) can produce quantisation artifacts around the enhanced highlights. This is reduced using a low pass filter only in the expanded regions, see Figure 6.

Finally, they ran a series of psychophysical experiments to determine the value of \(\rho\) for \(f\) using the DR-37P Dolby HDR monitor [Dol05]. The results showed that for outdoor scenes users preferred a high value of \(\rho\), which means a small percentage of dynamic range allocated to highlights, while for indoor scenes this was the contrary. For indoor and outdoor scenes of equal diffuse brightness users chose a low value for \(\rho\), so they preferred more range allocated to highlights. In conclusion from the analysis of the data, \(\rho = 0.66\) is a good general estimate.

This algorithm is designed for a specific task, the reproduction of highlights on HDR monitors. The use for other...
tasks, such as enhancement of videos, needs more processing and a classifier, which was underlined by authors’ evaluation experiment.

3.3.2. Enhancement of Bright Video Features for HDR Display

Didyk et al. [DMHS08] proposed an interactive system for enhancing brightness of LDR videos, targeting and showing results for DVD content. The main idea of the system is to classify a scene into three components: diffuse, reflections, and light sources, and then to enhance only reflections and light sources. The authors explained that diffuse components are difficult to enhance without creating visual artifacts and it was probably the intention of film-makers to show them saturated as opposed to light sources and clipped reflections. The system works on non-linear values, because the goal is the enhancement and non-physical accuracy.

The system consists of three main parts: pre-processing, classification, and enhancement of clipped regions, see Figure 7 for the pipeline. The pre-processing step generates data needed during the classification. In particular, it determines clipped regions using a flood-fill algorithm. At least one channel must be saturated (over 230 for DVD content), and luma values must be greater than 222. Also, in this stage optical flow is calculated as well as other features such as image statistics, geometric features and neighbourhood characteristics.

Classification determines lights, reflections, and diffuse regions in a frame and relies on a training set of 2,000 manually classified regions. Primarily, a support vector machine [Vap95] with kernel \( k(z, z') = \exp(-\gamma \|z - z'\|^2) \) performs an initial classification of regions. Subsequently, motion tracking improves the initial estimation, using a nearest neighbour classifier based on an Euclidean metric:

\[
d^2((z, x, t), (z', x', t')) = 50\|z - z'\|^2 + \|z - z'\|^2 + 5(t - t')^2
\]

where \( z \) are region features, \( x \) are coordinates in the image, and \( t \) is the frame number. This is allowed to reach a classification error of 12.6% on all regions used in the tests. Tracking of clipped regions using motion compensation further reduced the percentage of objects that require manual correction to 3%. Finally, the user can supervise classified regions, correcting wrong classifications using an intuitive user interface, see Figure 8.
Figure 6: The pipeline for the range expansion in Meylan et al.'s method [MDS07]. The original LDR image is expanded using Equation 1. Then, expanded luminance is filtered using a low pass filter. Finally, filtered expanded luminance and unfiltered one are linearly interpolated using a mask. This mask is calculated by thresholding LDR luminance with $\omega$. To remove noise, the mask is filtered with a dilatation and low pass filter.

Figure 7: The pipeline of the system proposed by Didyk et al. [DMHS08]: pre-processing (calculation of features vector, optical flow, and clipped regions), classification of regions using temporal coherence and a training set, user corrections (with updating of the training set), and brightness enhancement.
Clipped regions are enhanced by applying a non-linear adaptive tone curve, which is calculated based on partial derivatives within a clipped region stored in a histogram $H$. The tone curve is defined as an histogram equalization on the inverted values of $H$:

$$f(b) = k \sum_{j=2}^{b} (1 - H[j]) + t_2$$

where $t_2$ is the lowest luma value for a clipped region, $k$ is a scale factor that limits to the maximum boosting value $m$ (equal to 150\% for lights and 125\% for reflections):

$$k = \frac{m - t_2}{\sum_{j=1}^{N} (1 - H[j])}$$

where $N$ is the number of bins in $H$. To avoid contouring during boosting, the luma channel is filtered with bilateral filtering separating it into fine details and a base layer, which will be merged after luma expansion. See Section 3.6 for the validation part of this work. The method is semi-automatic, because intervention of the user is required.

### 3.4. Expand Map Models

The methods of Banterle et al. [BLDC06], its extensions [BLD\*07, BLDC08], and Rempel et al. [RTS\*07] use a guidance method to direct the expansion of the LDR content as opposed to global methods. Following the terminology used in Banterle et al. [BLDC06] we refer to these guidance methods as expand maps.

#### 3.4.1. Non-Linear Expansion using Expand Maps

A general framework for expanding LDR content for HDR monitors and IBL was proposed by Banterle et al. [BLDC06, BLD\*07]. The key points are the use of iTMO for expanding the range combined with a smooth field for the reconstruction of the lost over-exposed areas.

The first step of the framework is to linearise the input image, see Figure 9 for the pipeline. If the CRF is known, its inverse is applied to the signal. Otherwise, blind general methods can be employed such as Lin and et al.’s methods [LGYS04, LZ05]. Subsequently, the range of the image is expanded inverting a TMO. In their implementation, the inverse of the global Reinhard et al.’s operator [RSSF02] was used. This is because the operator has only two parameters, and range expansion can be controlled in a straightforward way. This iTMO is defined as:

$$L_w(x) = \frac{1}{2} L_{w, \text{Max}} L_{\text{white}} \left( L_d(x) - 1 + \sqrt{(1 - L_d(x))^2 + \frac{4}{L_{\text{white}}^2} L_d(x)} \right)$$

where $L_{w, \text{Max}}$ is the maximum output luminance in cd/m$^2$ of the expanded image, and $L_{\text{white}} \in (1, +\infty)$ is a parameter which determines the shape of the expansion curve. This is proportional to the contrast, authors suggested a value of $L_{\text{white}} \approx L_{w, \text{Max}}$ to increase the contrast while limiting artifacts due to expansion.

After range expansion, the expand map is computed. The expand map is a smooth field representing a low frequency version of the image in areas of high luminance. It has two main goals. The first is to reconstruct lost luminance profiles in over-exposed areas of the image. The second one is to attenuate quantisation or compression artifacts that can be enhanced during expansion. The expand map was implemented applying density estimation on samples generated using importance sampling (median-cut sampling [Deb05]). Finally, the expanded LDR image and the original one are combined using linear interpolation where the expand map acts as interpolation weight. Note that low luminance values are kept as in the original value. This avoids compression (for high $L_{\text{white}}$ values) for low values which can result in contouring.

![Figure 8: The interface used for adjusting classification results.](image)

![Figure 10: Application of Banterle et al.’s method [BLDC06, BLD\*07] for re-lighting synthetic objects: a) Lucy’s model is relighted using St. Peter’s HDR lightprobe. b) Lucy’s model is relighted using an expanded St. Peter’s LDR lightprobe (starting at exposure 0).](image)
The framework was extended for automatically processing images and videos in Banterle et al. [BLDC08]. This is achieved using 3D sampling algorithms, volume density estimation, edge transfers, and a number of heuristics for determining the parameters of each component of the framework. Moreover, a coloured expand map was adopted, allowing the reconstruction of clipped colours. The main problem is the speed, but real-time performances on high definition content can be achieved using point-based graphics on GPU.

The algorithm presents a general solution for visualisation on HDR monitors and IBL, see Figure 10 for an example. Moreover, it was tested using HDR-VDP [MDMS05] for both tasks to prove its efficiency compared with simple exposure methods. The main limit of the framework is that large over-exposed areas (more than 30% of the image) cannot be reconstructed using the expand map, producing grey smooth areas in the over-exposed areas. This is because there is not enough information to exploit.

### 3.4.2. LDR2HDR

A similar technique based on expand maps was proposed by Rempel et al. [RTS*07]. Their goal was real-time LDR expansion for videos. The algorithm pipeline is shown in Figure 11.

The first step of the LDR2HDR algorithm is to remove artifacts due to the compression algorithms of the media (such as MPEG) using a simple bilateral filter. Sophisticated artifact removal is not employed due to real-time constraints.

The next step of the method is to linearise the signal, using an inverse gamma function. Once the signal is linearised the contrast is stretched in an optimised way for the Dolby DR-37P HDR monitor [Dol05]. A simple linear contrast stretching is applied to boost values, however, they limited the maximum contrast to 5,000:1 to avoid artifacts. This means that the minimum value was mapped to 0.015 cd/m² while the maximum was mapped to 1,200 cd/m². To enhance brightness in bright regions a Brightness Enhance Function (BEF) is employed. This function is calculated applying a threshold of 0.92 (on a scale [0, 1] for LDR values). At this point the image is Gaussian filtered using a filter with a σ = 30 (150 pixels) which is chosen for 1920 × 1080 content. In order to increase contrast around edges an edge stopping function is used. Starting from saturated pixels, a flood-fill algorithm strategy is applied until an edge is reached, which is estimated using gradients. Subsequently, a morphological operator followed by a Gaussian filter with a smaller kernel is applied to remove noise. Finally, the BEF is mapped in the interval [1, α] where α = 4 and finally, is multiplied with the scaled image to generate the HDR image, see Figure 12 for an example. To improve efficiency the BEF is calculated using Laplacian pyramids [BA87], which is implemented on the GPU or FPGA [Dol05].

The algorithm was evaluated using HDR-VDP [MDMS05] comparing the linearised starting image with the generated HDR image. This evaluation was needed to show that the proposed method does not introduce spatial artifacts during expansion of the content. Note that LDR2HDR processes each frame separately which may be not temporally coherent due to the nature of the BEF.

### 3.5. User Based Models

Since it may not always be possible to recover missing HDR content using automatic approaches, a different, user-based approach was proposed by Wang et al. [WWZ*07], whereby detailed HDR content can be added to areas that are meant to be expanded.

The authors demonstrated the benefits of an in-painting system to recover lost details in over-exposed and under-
exposed regions of the image, combined with a luminance boosting. The whole process was termed hallucination, and their system presents a mixture between automatic and user-based approaches.

The first step of hallucination is to linearise the signal, see Figure 13 for the complete pipeline. This is achieved with an inverse gamma function with $\gamma = 2^2$, which is the standard value for DVDs and television formats [ITU90]. After this step, the image is decomposed into large scale illumination and fine texture details. This is achieved by applying bilateral filtering to the image $I_{ob}$ obtaining a filtered version $I_{lf}$.

The texture details are obtained as $I_d = I/I_{lf}$. Radiance for large scale illumination $I_{lf}$ is estimated using a linear interpolation of elliptical Gaussian kernels. Firstly, a weight map, $w(x)$, is calculated for each pixel:

$$w(x) = \begin{cases} \frac{C_{ue} - Y(x)}{C_{ue}} & Y(x) \in [0,C_{ue}) \\ 0 & Y(x) \in [C_{ue},C_{oe}) \\ \frac{Y(x) - C_{oe}}{1 - C_{oe}} & Y(x) \in [C_{oe},1] \end{cases}$$

where $Y(x) = R_s(x) + 2G_s(x) + B_s(x)$, and $C_{ue}$ and $C_{oe}$ are respectively the thresholds for under-exposed and over-exposed pixels. The authors suggested values of 0.05 and 0.85 for $C_{ue}$ and $C_{oe}$ respectively. Secondly, each over-exposed region is segmented and fitted with an elliptical Gaussian lobe $G$, where variance of the axis is estimated using region extents, and the profile is calculated using an optimisation procedure based on non over-exposed pixels at the edge of the region. The luminance is blended using a simple linear interpolation:

$$O(x) = w(x)G(x) + (1 - w(x))\log_{10}Y(x)$$

Optionally users can add Gaussian lobes using a brush.

The texture details $I_d$ are reconstructed using a texture synthesis technique similar to [BVSO03], where the user can select an area as a source region by drawing it with a brush. This automatic synthesis has some limits when scene understanding is needed, therefore a warping tool is included. This allows the user to select with a stroke-based interface a source region and a target region, and pixels will be transferred. This is a tool similar to the stamp and healing tools in Adobe Photoshop [Ado07].

Finally, the HDR image is built blending the detail and the large scale illumination, this is performed using Poisson image editing [PGB03] in order to avoid seams in the transition between expanded over-exposed areas and well-exposed areas.

This system can be used for both IBL and visualisation of images, and compared with other algorithms it may maintain details in clamped regions. However, the main problem of this approach is that it is user based and not automatic, which potentially limits its use to single images and not videos.

### 3.6. Validation

The development of methods for LDR expansion has produced various algorithms with different features. Therefore,
there is a need to determine the quality performances of such algorithms to understand which method is more suitable for a given image and task. Moreover, the analysis of performances can help to highlight important features, such non-linearity, that can be important for the design of future expansion techniques.

3.6.1. HDR-VDP Comparisons

Banterle et al. [BLDC06, BLD*07] used the HDR Visual Difference Predictor (HDR-VDP) [MDMS05] for validating the quality of reconstruction against a ground truth and simple expansion operator without expand maps. The results showed that their proposed method reconstructs closer to the reference the missing HDR content.

Moreover, HDR-VDP was applied to compare re-lighted images with an HDR reference. This showed that LDR expansion allows a small error in the case of IBL.

3.6.2. Pairwise Comparisons Study for Video Sequences

In Section 3.3.2, Didyk et al. [DMHS08] presented a new operator for the expansion of LDR videos based on classification. Enhanced videos generated with this method were compared with the original videos and the ones generated, using the only method suitable for videos at the time, Rempel et al.’s [RTS*07] LDR2HDR. Comparisons were performed running a paired comparisons psychophysical experiment [Dav88, LCTS05] with 14 naïve subjects using a LCD Barco Coronis Color 3MP Diagnostic Luminance (12-bit per colour channel). The participants ranked 9 videos times 3 combinations: original video, Rempel et al. and their method. The study was analysed using a similar approach to Ledda et al. [OA07]. The experiment showed that for the overall scores Didyk et al.’s method was preferred with statistical significance compared to both the original video and the Rempel et al.’s one. However, there was no statistical significance between this method and other ones for six of the considered videos.

3.6.3. Dynamic Range Independent Image Quality Assessment

Aydin et al. [AMMS08] proposed a new perceptual metric (DI-IQA) which allows the comparison of images independently from their dynamic range. This metric can detect the loss and amplification of contrast, and the change of structure in images.

Due to the capabilities of this metric, quantisation artifacts and changes in the image details visibility can be quantified where they happen. Therefore, it can be employed to validate the quality of expansion algorithms avoiding time consuming psychophysical experiments. Authors presented a few examples where they applied the metric to expanded images showing when the signal starts to be distorted in function of the expansion, see Figure 14 for an example.

3.6.4. Pairwise Comparisons Studies for Image Visualisation and Image Based Lighting

Banterle et al. [BLD*08] proposed a psychophysical study for the evaluation of expansion algorithms based on pairwise comparisons methodology [Dav88, LCTS05] using an HDR reference image displayed on the Dolby DR-37p HDR monitor [Dol05]. The study involved 24 participants, and five algorithms were tested: Banterle et al. [BLDC06, BLD*07] (B), Meylan et al. [MDS06, MDS07] (M), Wang et al. [WWZ*07] (W), Rempel et al. [RTS*07] (R), and Akyüz

Figure 13: The pipeline of the Wang et al. method [WWZ*07].
simple techniques that apply single or multiple linear scale expansions, such as A and M. The more computationally expensive methods B, R and W, are better at recreating HDR than simple methods. Even if a linear scale can elicit an HDR experience in an observer, as shown in [AFR’07], it does not correctly reproduce the perception of the original HDR image.

3.6.5. Exposure Understanding for Content Expansion

Martin et al. [MFS’08] presented an ongoing psychophysical study on evaluation of expansion algorithms. This study is divided in two parts. The first part of the study is focused on perception of exposure in images. The results of the first part showed that high-level semantics are needed for a proper classification of exposure. Moreover, an asymmetry in the perception of under-exposed and over-exposed images was found.

The second part consisted of side by side evaluation of the following expansion methods on a Dolby DR-37p monitor [Dol05] with images at different exposure levels: original LDR, Banterle et al. [BLDC06, BLD’07]. Rempel et al. [RTS’07], and Akýüz [AFR’07]. The results of this experiment have not currently been published.

3.7. Overview

In Table 1 we present an overview of all the methods discussed in this section, summarising what techniques they use, and how they compare in terms of quality and performance. We find that most methods expand the dynamic range using either a linear or non-linear function, while Meylan et al. use a two-scale linear function. The reconstruction methods aim at smoothly expanding the dynamic range and a variety of methods are proposed. Unsurprisingly, the choice of expansion method and reconstruction influences the computational performance of the method and the quality. We present performance based on the timings from the individual papers and/or the complexity of the computation involved, where fast performance would make it possible to perform in real-time on current hardware while slow would require a handful of seconds. Wang et al.’s method requires a manual intervention somewhat hindering real-time performance. The quality results we present are based in other publications, primarily the psychophysical experiments shown in Banterle et al. [BLD’08]. It is clear that different methods are suitable for different applications, and the more straightforward methods are faster and more suitable for IBL or just improving highlights. For more complex still scenes and/or videos where further detail may be desirable, the more complex expansion methods may be preferable.

4. HDR Compression using Tone Mapping and Inverse Tone Mapping

HDR expansion methods have not only been employed for the generation of content from a single exposure image, but
they have proven beneficial for HDR content compression. These methods typically comprise of the compression of the dynamic range via tone mapping. The tone mapped image is subsequently encoded via traditional compression methods such as JPEG, in the case of images, or MPEG in the case of videos. These two steps comprise the encoding aspect of the compression. Decoding takes the role of the LDR compression’s decoding method followed by an HDR expansion, usually inverting the method that was used for the dynamic range compression.

This approach to the compression of HDR content has the advantage of re-using previous compression schemes and standards. Also, it can allow backward-compatibility because the function for HDR expansion can be easily stored in an extension header of a standard. These functions require only a few parameters to be stored.

4.1. Backward Compatible JPEG-HDR

JPEG-HDR is an extension to the JPEG compression scheme for HDR images by Ward and Simmons [WS04, WS05]. The method does not use an explicit iTMO, nevertheless a spatial inverse function called Ratio Image (RI) is employed.

The encoding, see Figure 15, starts with the tone mapping of the HDR image discretised to 8-bit. After this, the original HDR image is divided by the tone mapped one obtaining the RI which is stored as a sub-band. The RI can be down-sampled reducing the sub-band size, because the HVS has a limited ability to detect large and high frequency changes in luminance. This fact was also exploited in Setzeen et al. [SHS’04] to improve the efficiency of HDR displays. However, down-sampling needs correction of the image, because the naive multiplication of a down-sampled image times the tone mapped LDR image can produce halos/glare around edges. This problem can be solved in two ways: pre-correct and post-correct. The former method introduces corrections in the tone mapped image. This is achieved by down-sampling and afterwards up-sampling the RI image obtaining RI_L. Subsequently, the original HDR image is divided by RI_L, which is a tone mapped image with corrections. While this approach is effective, it can produce artifacts in the LDR image for the backward compatibility and this cannot be acceptable in many applications. The latter method consists of an up-sampling with guidance which is more expensive than the pre-correction one. While RI_d is discretised at 8-bit in the logarithmic domain and stored in application markers of JPEG, the tone mapped layer needs further processing for preserving colours. Two techniques are employed to solve this problem: compression of the gamut and a new YC_bC_r encoding. The gamut compression produces a global desaturation. Given the following definition of saturation:

\[
S(x) = 1 - \frac{\min(R(x), G(x), B(x))}{L_w(x)}
\]

the desaturation of each colour channel is achieved by:

\[
\begin{align*}
R_c(x) &= \left(1 - S(x)^\beta\right)^{\alpha} L_w(x) + S(x)^{\beta} R_c(x) \\
B_c(x) &= \left(1 - S(x)^\beta\right)^{\alpha} L_w(x) + S(x)^{\beta} B_c(x)
\end{align*}
\]

where \(\alpha \leq 1\) is a parameter which controls the level of saturation kept during colour encoding, \(\beta\) is a parameter which determines the colour contrast, and \(S(x) = \alpha S(x)^{\beta - 1}\) is the desaturation level. After this step, the image is encoded in a modified YC_bC_r colour space, because it has a larger gamut than RGB colour space. Therefore, unused YC_bC_r values can be exploited to preserve the original gamut of an HDR image. This is achieved by the following mapping:
Xu et al. [XPH05] proposed a simple pre-processing technique which enables the JPEG 2000 standard [CSE00] to encode HDR images. The main idea is to transform floating point data in unsigned short integers (16-bit), that are supported by JPEG 2000 standard.

The encoding phase starts with the reduction of the dynamic range by applying a logarithm to the RGB values:

\[
\begin{align*}
R'_w(x) &= \log R_w(x) \\
G'_w(x) &= \log G_w(x) \\
B'_w(x) &= \log B_w(x)
\end{align*}
\]

Subsequently, the floating point values in the logarithm domain are discretised to unsigned short integers:

\[
\begin{align*}
\left[ \frac{R_w(x)}{2^7 - 1} \right] &\geq x_c \\
\left[ \frac{G_w(x)}{2^7 - 1} \right] &\geq x_c \\
\left[ \frac{B_w(x)}{2^7 - 1} \right] &\geq x_c
\end{align*}
\]

where \(x_c\) is a quality setting between 0.6-3.75 bpp for quality settings between 57 - 99%. However, quality degrades rapidly for JPEG quality below 60%, but only 2.5% of pixels were visibly different with a quality set at 90%, and only 0.1% with maximum quality.

Most importantly, the method is backward compatible, because RLI is encoded using only extra application markers of JPEG. When an old application or one that is not designed for HDR imaging will open a JPEG-HDR file, it will display only the tone mapped layer allowing the user to have access to the HDR part of the content.

\[R'(x) = \begin{cases} 
1.055R(x)^{0.42} - 0.055 & \text{if } R(x) > t_r \\
12.92R(x) & \text{if } |R(x)| \leq t_r \\
-1.055(-R(x))^{0.42} + 0.055 & \text{if } R(x) < -t_r
\end{cases} \]

which is repeated for the green and blue channel. Finally, the standard mapping from RGB to YC_bC_r is used for the JPEG encoding.

The decoding for the pre-correction case consists of few steps, see Figure 16 for the complete pipeline. Firstly, the tone mapped layer is decoded using a JPEG decoder and the gamut is expanded inverting Equation 2. After this step, the RLI image is decoded, expanded (from logarithmic domain to linear domain), and up-sampled to the resolution of the tone mapped layer. Finally, the image is recovered by multiplying the tone mapped layer by the RLI image.

A first study [WS04] was conducted to determine a good TMO for compression purposes, which was based on comparison with the original HDR images using VDP [Dal93]. In this experiment different TMOs were compared such as histogram adjustment [LRP97], global photographic tone reproduction operator [RSSF02], fast bilateral filtering operator [DD02] and the gradient operator [FLW02]. Experiments showed that the fast bilateral filtering operator performed the best followed by the global photographic tone reproduction one. A second study was carried out to test image quality and compression rates on a data set of 217 HDR images. The data set was compressed using JPEG-HDR at different quality settings using the global photographic operator, RGBE, OpenEXR and LogLuv TIFF to study compression rates. HDR images compressed using JPEG-HDR were compared with original ones using VDP to quantify the quality of the resultant images. The study showed that the method can achieve a compression rate between 0.6-3.75 bpp for quality settings between 57 - 99%. However, quality degrades rapidly for JPEG quality below 60%, but only 2.5% of pixels were visibly different with a quality set at 90%, and only 0.1% with maximum quality.

Most importantly, the method is backward compatible, because RLI is encoded using only extra application markers of JPEG. When an old application or one that is not designed for HDR imaging will open a JPEG-HDR file, it will display only the tone mapped layer allowing the user to have access to the HDR part of the content.

4.2. HDR-JPEG 2000

The method using JPEG 2000 lossy mode was compared to JPEG-HDR [WS05] and HDRV [MKMS04], and when using JPEG 2000 lossless mode it was compared with RGBE [War91], LogLuv [Lar98], and OpenEXR [Ind02]. The metrics used for the comparison were RMSE in the logarithm domain and Lubin’s VDP [Lub95]. The results of these comparisons showed that HDR-JPEG 2000 in lossy mode is superior to JPEG-HDR and HDRV, especially at low bit rates when these methods produce visible artifacts. Nevertheless, the method does not perform well when lossless JPEG 2000 is used, because the file size is higher than RGBE, LogLuv, and OpenEXR (these methods are lossy in terms of per-pixel float precision, but not spatially over pixel neighbourhoods).

Finally, values are exponentiated to get the final dynamic range:

\[
\begin{align*}
R_w(x) &= e^{R'_w(x)} \\
G_w(x) &= e^{G'_w(x)} \\
B_w(x) &= e^{B'_w(x)}
\end{align*}
\]

The HDR-JPEG 2000 algorithm is a straightforward method for lossy compression of HDR images at high quality, without artifacts at low bit rates. However, the method...
Figure 15: The encoding pipeline for JPEG-HDR for pre-correction case by Ward and Simmons [WS04, WS05].

Figure 16: The decoding pipeline for JPEG-HDR by Ward and Simmons [WS04, WS05].
is not suitable for real-time graphics, because fixed time look-ups are needed. Also, the method does not exploit all the compression capabilities of JPEG 2000 because it operates at high level. For example, separate processing for luminance and chromaticity could reduce the size of the final image while keeping the same quality.

### 4.3. Compression and Companding High Dynamic Range Images with Sub-bands Architectures

Li et al. [LSA05] presented a general framework for tone mapping and inverse tone mapping of HDR images based on multi-scale decomposition. While the main goal of the algorithm is tone mapping, in addition, the framework can also compress HDR images. A multi-scale decomposition splits a signal $s(x)$ (1D in this case) into $n$ sub-bands $b_1(x), ..., b_n(x)$ with $n$ filters $f_1, ..., f_n$, in a way the signal can be reconstructed as:

$$s(x) = \sum_{i=1}^{n} b_i(x)$$

Wavelets [SDS95] and Laplacian pyramids [BA87] are examples of multi-scale decomposition that can be used in Li et al.’s framework.

![Figure 17: An example of tone mapping using the multi-scale decomposition with Haar Wavelets. a) Activity map, b) Gain map and c) Tone mapped luminance.](image)

The main concept is to apply a gain control to each sub-band of the image to compress the range. For example, a sigmoid expands low values and flats peaks, however it introduces distortions that can appear in the final reconstructed signal. In order to avoid such distortions, a smooth gain map inspired by neurons was proposed. The first step, is to build an activity map, reflecting the fact that the gain of a neuron is controlled by the level of its neighbours. The activity map is defined as:

$$A_i(x) = G(\sigma_i) \otimes |B_i(x)|$$

where $G(\sigma_i)$ is a Gaussian kernel with $\sigma_i = 2^i \sigma_1$ which is proportional to $i$, the sub-band’s scale. The activity map is used to calculate the gain map, which turns gain down where activity is high and vice versa:

$$G_i(x) = p(A_i(x)) = \left( \frac{A_i(x) + \varepsilon}{\delta} \right)^{\gamma - 1}$$

where $\gamma \in [0, 1]$ is a compression factor, and $\varepsilon$ is the noise level that prevents the noise from being seen. $\delta = \alpha \sum_{x} A_i(x)/(M)$ is the gain control stability level where $M$ is the number of pixels in the image, $\alpha_i \in [0.1, 1]$ is a constant related to spatial frequency. Once the gain maps are calculated, sub-bands can be modified:

$$B'_i(x) = G_i(x)B_i(x)$$  \hspace{1cm} (4)

Note that it is possible to calculate a single activity map for all sub-bands by pooling all activity maps:

$$A_{ag}(x) = \sum_{i=1}^{n} A_i(x)$$

From $A_{ag}$, a single gain map $G_{ag} = p(A_{ag})$ is calculated for modifying all sub-bands. The tone mapped image is finally obtained summing all modified sub-bands $B'_i$, see Figure 17. The compression is applied only to the V channel of an image in the HSV colour space. Finally, to avoid oversaturated images $S$ can be reduced by $\alpha \in [0.5, 1]$. The authors presented a comparison with the fast bilateral filter operator [DD02] and gradient domain operator [FLW02].

The framework can be additionally used for the compression task, applying expansion after compression, called companding. The expansion operation is obtained by a straightforward modification of Equation 4:

$$B'_i(x) = \frac{B_i(x)}{G_i(x)}$$

A straightforward companding operation is not sufficient for compression especially if the tone mapped image is compressed using lossy codecs. Therefore, the companding operation needs to be iterative to determine the best values for the gain map, see Figure 18. The authors proposed to compress the tone mapped image using JPEG. In this case a high bit-rate is needed (1.5 bpp - 4bpp) with chrominance sub-sampling disabled to avoid that JPEG artifacts are amplified during expansion, because a simple up-sampling strategy is adopted.

### 4.4. Backward Compatible HDR-MPEG

Backward compatible HDR-MPEG is a codec for HDR videos that was introduced by Mantiuk et al. [MEMS06].
As in the case of JPEG-HDR this algorithm is an extension to the standard MPEG-4 codec (H.264) [WSBL03] that works on top of the standard encoding/decoding stage allowing backward compatibility. In a similar way to JPEG-HDR each frame is divided into an LDR part, using tone mapping, and an HDR part. However, in this method, the reconstruction function (RF) a tabled iTMO is employed instead of a RI. HDR-MPEG is a natural extension of perception motivated video encoding (HDRV) [MKMS04]. However, the primary features of the HDRV codec design is that it is a modification of standard MPEG-4 with new steps in the encoding/decoding stage such as the perceptual luminance encoding. Moreover, HDRV was designed for a target of 10-11 bit for luminance, a format that is rarely supported in software and hardware, which compromises its backward compatibility.

The decoding stage takes as input an HDR video in the XYZ colour space and it applies tone mapping to each HDR frame obtaining LDR frames as a first step, see Figure 19 for the complete pipeline. These are coded with MPEG-4, stored in an LDR stream, and finally decoded to obtain a uncompressed and MPEG quantised frames. Subsequently, the LDR frame and the HDR frame are converted to a common colour space. For both HDR and LDR frames CIE 1976 Uniform Chromaticity \((u', v')\) coordinates are used to code chroma. While non-linear luma of sRGB is used for LDR pixels, a different luma coding is used because sRGB non-linearity is not suitable for high luminance ranges \([10^{-5}, 10^{10}]\), see [MEMS06]. This luma coding, at 12-bit, for HDR luminance values is given as:

\[
l_w = f(l_w) = \begin{cases} 
209.16 \log(L_w) - 731.28 & \text{if } L_w \geq 10469 \\
826.81 L_w^{0.10013} - 884.17 & \text{if } 5.6046 \leq L_w < 10469 \\
17.554 L_w & \text{if } L_w < 5.6046 
\end{cases}
\]

where its inverse transform, \(g(l_w) = f^{-1}(l_w)\), is:

\[
L_w = g(l_w) = \begin{cases} 
32.994 \exp(0.0047811 l_w) & \text{if } l_w \geq 1204.7 \\
7.3014 e - 30(l_w + 884.17)^{9.987} & \text{if } 98.381 \leq l_w < 1204.7 \\
0.056968 l_w & \text{if } l_w < 98.381 
\end{cases}
\]

At this point both the HDR and the LDR frames are in a comparable colour space, and an RF, that maps \(l_d\) to \(l_w\), is calculated in a straightforward way by averaging \(l_w\), which falls into one of 256 bins representing the \(l_d\) values:

\[ \text{Figure 18: The optimisation companding pipeline of Li et al. [LSA05].} \]
Figure 19: The encoding pipeline for Backward Compatible HDR-MPEG by Mantiuk et al. [MEMS06].

RF(i) = \frac{1}{|\lambda|} \sum_{x \in \lambda_l} l_w(x) \quad \text{where} \quad \lambda_l = \{i|l_d(x) = l\}

where \( I \in [0, 255] \) is an index of a bin, \( l_d(x) \) and \( l_w(x) \) are respectively the luma for LDR and HDR pixel at \( x \). RF for chromaticity is approximated imposing \((u'_d, v'_d) = (u'_w, v'_w)\).

Once RFs are calculated for all frames, they are stored in an auxiliary stream using Huffman encoding.

After this stage a residual image is calculated for improving overall quality:

\[ r_l(x) = l_w(x) - RF(l_d(x)) \]

The residual image is discretised at 8 bit, using a quantisation factor different for each bin based on its maximum magnitude value, which leads to:

\[ \hat{r}_l(x) = \frac{r_l(x)}{q(m)} \times 127 \quad \text{where} \quad m = k \Leftrightarrow i \subset \lambda_k \]

where \( q(m) \) is the quantisation factor which is calculated for a bin \( \lambda_k \) as:

\[ q(m) = \max q_{\min} \left( \frac{\max_{x \in \lambda_k} |r_l(x)|}{127} \right) \]

\( \hat{r}_l \) needs to be compressed in a stream using MPEG, but a naïve compression would generate a low compression rate, because a large amount of high frequencies are present in \( \hat{r}_l \). In order to improve the compression rate, the image is filtered removing frequencies in regions that are not distinguishable by the HVS. This is achieved by using the original HDR frame as guidance to the filtering. The filtering is performed in the wavelet domain, and it is applied only to the three finest scales modeling contrast masking, and lower sensibility to high frequencies.

The decoding stage is quite straightforward. MPEG streams (tone mapped video and residuals) and RF streams are decoded, see Figure 20 for the complete pipeline. Then, an HDR frame is reconstructed applying firstly its RF to the LDR decoded frame, and secondly adding residuals to the expanded LDR frame. Finally, CIE Luv values are converted to XYZ ones using Equation 5 for luminance.

HDR-MPEG was evaluated using three different metrics: HDR VDP [MDMOS05], universal image quality index (UQI) [WB02], and classic Signal to Noise Ratio (SNR).

As in the case of JPEG-HDR, there was first a study that explored the influence of a TMO on quality/bit-rate. This experiment was performed using different TMOs such as time dependent visual adaption [PTYG00], fast bilateral filtering [DD02], photographic tone reproduction [RSSP02], gradient domain [FLW02] and adaptive logarithmic mapping [DMAC03]. These TMOs were modified to avoid temporal flickering and applied to a stream using default parameters. The study showed that most of these TMOs have the same performances except the gradient domain one, which creates larger streams than others. However, this TMO generated more attractive images for backward compatibility, therefore the choice of a TMO for the video compression de-
Figure 20: The decoding pipeline for Backward Compatible HDR-MPEG by Mantiuk et al. [MEMS06].

4.5. Encoding of High Dynamic Range Video with a Model of Human Cones

Similarly to Li et al. [LSA05], Van Hateren [Hat06] proposed a new TMO based on a model of human cones [HL06] which can be inverted to encode HDR images and videos. The TMO and iTMO work in troland units (td), the measure of retinal illuminance \( I \), which is derived by the scene luminance in cd/m\(^2\) multiplied by the pupil area in mm\(^2\). Van Hateren proposed a temporal and a static version of its TMO.

The temporal TMO is designed for HDR videos and presents low-pass temporal filters for removing photon and source noise, see Figure 21. The TMO starts by simulating the absorption of \( I \) by visual pigment, which is modelled by two low-pass temporal filters that are described in terms of a differential equation:

\[
\frac{dy}{dt} + \frac{1}{\tau} y = \frac{1}{\tau} x
\]

where \( \tau \) is a time constant, \( x(t) \) is the input at time \( t \), and \( y(t) \) is the output. At this point, a strong non linearity is applied to the result of low-pass filters \( E^* \) for simulating the breakdown of cyclic guanosine monophosphate (cGMP) by enzymes (cGMP is a nucleotide, that controls the current across the cell membranes):

\[
\alpha = \frac{1}{\beta} = \left( c_{\beta} + k_{\beta}E^* \right)^{-1}
\]

where \( k_{\beta}E^* \) is the light-dependent activity of an enzyme, and \( c_{\beta} \) the residual activity. The breakdown of cGMP is counteracted by the production of cGMP: a highly non-linear feedback loop under control of inter-cellular calcium. This system is modelled by a filtering loop which outputs the current across cell membrane, \( I_{os} \) (the final tone mapped value), by the outer segment of a cone.

Van Hateren showed that the range expansion is quite straightforward by inverting the feedback loop. However, the process cannot be fully inverted because the first two low-pass filters are difficult to invert, so this results in \( I' \approx I \). In order to fully invert the process for inverse tone mapping purposes Van Hateren proposed a steady version of the TMO, a global TMO, defined as:

\[
I_{os} = \frac{1}{\left(1 + (a_C I_{os})^4\right)(c_{\beta} + k_{\beta}I)}
\]  \hspace{1cm} (6)

where \( a_C \) is a scaling constant. Equation 6 can be easily inverted as:
Van Hateren applied its TMO and iTMO to uncalibrated HDR images, which were scaled by a harmonic mean. The results showed that the method does not need gamma correction, removes noise, and accounts for light adaptation, see Figure 22. The main drawbacks of the TMO scheme is designed to be backward compatible. The main differences from HDR-MPEG are the presence of a min-isation step for optimising tone mapping parameters, the differences from HDR-MPEG is performed in the logarithm domain to match HVS perception and avoid outliers at high values. The error to minimise is given by:

$$E(I) = \sum_{x\in I} (\log(L_w(x)) - \log(g(L_d(x))))^2$$

(7)

for determining $n$ and $k$. The optimum solution is uniquely determined imposing the partial derivatives of $E$ for $k$ and $n$ equal to zero, leading to:

$$k = \exp\left(\frac{\sum_x B(x)^2 \left(\sum_x A(x) - \sum_x B(x)\right) \left(\sum_x A(x) B(x)\right)}{M \left(\sum_x B(x)^2\right) - \left(\sum_x B(x)^2\right)^2}\right)$$

$$n = \frac{M \left(\sum_x B(x)^2\right) - \left(\sum_x B(x)^2\right)^2}{M \left(\sum_x A(x) B(x)\right) - \left(\sum_x A(x) \sum_x B(x)\right)}$$

where $M$ is the number of pixels, and $A$ and $B$ are defined as:

$$A(x) = \log L_w(x) \quad B(x) = \log \left(\frac{L_d(x)}{1 - L_d(x)}\right)$$

Once the parameters are determined the image is tone mapped and encoded using JPEG. Then residuals are calculated to improve quality. They are calculated as:

$$R(x) = \left(\frac{L_w(x)}{g(L_d(x)) + \varepsilon}\right)^\gamma$$

where $\gamma \in (0, 1]$ is a constant, and $\varepsilon$ is a small value to avoid
Figure 23: The encoding pipeline presented in Okuda and Adami [OA07].

Figure 24: The decoding pipeline presented in Okuda and Adami [OA07].
discontinuities chosen by the user. Finally, $R$ is encoded using a wavelet image compression scheme.

The decoding stage is straightforward. Once the LDR image and residuals are decoded using a JPEG decoder and a wavelet decoder, final HDR values are recovered by:

$$L_w(x) = R(x) \cdot \left( g(L_d(x)) + \varepsilon \right)$$

Two colour compression methods are presented to preserve distortions caused by tone mapping. The first one is a modification of Ward and Simmons [WS05] where $\alpha$ and $\beta$ are calculated with a quadratic minimisation using an error function similar to Equation 7. The second method is to apply a polynomial $P(x)$ for each LDR colour channel, assuming that a polynomial relationship exists between LDR and HDR values. Coefficients of $P(x)$ are fitted using the Gaussian weighted difference between the original and the reconstructed HDR channels.

The compression scheme was evaluated on a data set of 12 HDR images and compared with JPEG-HDR and HDR-MPEG using two metrics. The first one is the mean square error (MSE) in CIELAB colour space [Fat05] to test overall quality:

$$MSE = \frac{1}{M} \sum_{x \in \Omega} \Delta_{\text{Lab}}(x)$$

$$\Delta_{\text{Lab}}(x) = \left( L_{w,1}(x) - L_{w,2}(x) \right)^2 + \left( a_1(x) - a_2(x) \right)^2 + \left( b_1(x) - b_2(x) \right)^2$$

The second metric is MSE in the Daly’s non-linearity domain [Dal93] to test reconstructed luminance:

$$L_{\text{MSE}}(I_1, I_2) = \frac{1}{M} \sum_{x \in \Omega} \left( Dn(L_{w,1}(x)) - Dn(L_{w,2}(x)) \right)^2$$

$$Dn(x) = \frac{x}{x + 12 \cdot 0.65}$$

In their experiments the proposed method achieved better results for both metrics in comparison with JPEG-HDR and HDR-MPEG at different bit rates. While the quality of this method is up to two times better than HDR-MPEG and JPEG-HDR at high bit rates (around 8-10 bits), it is comparable for low bit rates (around 1-4 bits).

4.7. HDR Texture Compression using Inverse Tone Mapping

A similar compression scheme to Okuda and Adami [OA07] was presented by Banterle et al. [BDLC08] for HDR texture compression. This method was designed to take advantage of graphics hardware. The generalised framework presented uses a minimisation process that takes into account the compression scheme for tone mapped images and residuals. Moreover, it was shown that up-sampling of tone mapped values before expansion does not introduce visible errors.

Authors employed the global Reinhard et al. operator [RSSP02] and its inverse [BLDC06] in their implementation. The forward operator is defined as:

$$f(L_d(x)) = L_d(x) = \frac{\alpha L_w(x) + b_1(x) + b_2(x)}{L_{w,11}(x) + L_{w,11}}$$

$$[R_d(x), G_d(x), B_d(x)]^T = \frac{L_d(x) + w(x) \cdot [R_w(x), G_w(x), B_w(x)]}{L_{w,\text{white}}}$$

where $L_{w,\text{white}}$ is the luminance white point, $L_{w,11}$ is the harmonic mean, and $\alpha$ is the scale factor. While the inverse is given by:

$$g(L_d(x)) = f^{-1}(L_d(x)) = L_w(x) = \frac{L_{w,11} \cdot L_{w,11}}{2 \cdot \alpha} \left( L_d(x) - 1 + \sqrt{(1 - L_d(x))^2 + 4L_d(x)} \right)$$

$$[R_w(x), G_w(x), B_w(x)]^T = \frac{L_{w,\text{white}} \cdot [R_d(x), G_d(x), B_d(x)]}{L_d(x)}$$

The first stage of encoding is to estimate parameters of the TMO, similarly to [Ref02], and to apply a colour transformation, see Figure 25 for the encoding pipeline. However, this last step can be skipped because S3TC does not support colour spaces with separated luminance and chromaticity. Subsequently, the HDR texture and estimated values are used as input in a Levenberg-Marquadt minimisation loop which ends when the local optimum for TMO parameters is reached. In the loop, the HDR texture is firstly tone mapped and encoded with S3TC. Secondly, residuals are calculated and encoded using S3TC. Finally, the image is reconstructed, and error is calculated and new TMO parameters are estimated. When local optimum is reached, the HDR texture is tone mapped with these parameters and encoded using S3TC with residuals in the alpha channel.

The decoding stage is straightforward and can be implemented in a simple shader on GPU, see Figure 26 for the decoding pipeline. When a texel is needed in a shader, the tone mapped texture is fetched and its luminance is calculated. The inverse tone mapping, uses these luminance values, combined with the TMO parameters, to obtain the expanded values which are then added to the residuals. Finally, luminance and colours are recombined. Note that the inverse operator can be pre-computed into a 1D texture to speed-up the decoding. Moreover, computations can be sped-up applying filtering during the fetch of the tone mapped texture. This is because the filtering is applied to coefficients of a polynomial function. Authors proposed a bound of this error, showing that is not significant in many cases.

This proposed scheme was compared to RGBE [War91],
Figure 25: The encoding pipeline presented in Banterle et al. [BDLC08].

Figure 26: The decoding pipeline presented in Banterle et al. [BDLC08].

Figure 27: A comparison of real-time decoding schemes on current graphics hardware applied to St. Peter’s Cathedral lightprobe: a) Banterle et al.’s scheme [OA07]. b) Wang et al.’s one [WWZ*07] showing visible contouring artifacts.

Munkberg et al.’s method [MCHAM06], Roimela et al.’s scheme [RAI06], and Wang et al.’s scheme [WWZ‘07] using HDR-VDP [MDMS05], mPSNR [MCHAM06], and RMSE in the logarithm domain [XPH05]. The results showed that the new schemes presents a good trade-off between quality and compression, as well as the ability to decode textures in real-time. Moreover, it has a better quality on average than Wang et al.’s method, the other real-time decoding scheme, avoiding contouring artifacts, see Figure 27. The main disadvantage of this method is not being able to efficiently encode the luminance and chromaticity due to limits of S3TC.

4.8. Validation

The evaluation of quality for image compression schemes is usually performed using image metrics such as: HDR Visual Difference Predictor (HDR-VDP) [MDMS05], a perceptual metric, Root Mean Squared Error (RMSE) in the
log₂[RGB] domain [XPH05], and multi-exposure Peak Signal Noise Ratio (mPSNR).

4.8.1. Root Mean Squared Error in the log₂[RGB] domain

The RMSE in the log₂[RGB] domain was proposed by Xu et al. [XPH05], which is defined as follows:

$$\text{RMSE}(I, \hat{I}) = \sqrt{\frac{1}{n} \sum \Delta_{\text{RGB}}(x)}$$

where $I$ is the reference image and $(R,G,B)$ its red, green and blue channels, $\hat{I}$ the comparison image and $(\hat{R}, \hat{G}, \hat{B})$ its channels, $n$ the number of pixels of the image. A small RMSE value means that image $\hat{I}$ is close to the reference, zero means that they are the same, while a high value means that they are very different.

4.8.2. mPSNR

mPSNR is an extension of PSNR metric to HDR domain by Munkberg et al. [MCHAM06]. This takes a series of exposures which are tone mapped using a simple gamma curve:

$$T(X, c) = \left[ \frac{\min(255 \times x)^{\gamma}}{255} \right]$$

where $c$ is the current f-stop, $X$ is a colour channel, and $\gamma = 2.2$. Then the classic Mean Square Error (MSE) is computed:

$$\text{MSE}(I, \hat{I}) = \frac{1}{n \times p} \sum_{c=p_{\text{Min}}}^{p_{\text{Max}}} \sum_{x} \Delta_{\text{RGB}}^2(x) = \sum_{x} \Delta R_c^2(x) + \Delta G_c^2(x) + \Delta B_c^2(x)$$

where $p_{\text{Min}}$ and $p_{\text{Max}}$ are respectively the minimum and maximum exposures, $p$ is the number of used exposures, $n$ is the number of pixels in the image, $\Delta R_c(x) = T(R(x), c) - T(\hat{R}(x), c)$ for the red colour channel, and similarly for green and blue channels. Finally, the m-SNR is calculated using the standard formula:

$$\text{mPSNR}(I, \hat{I}) = 10 \log_{10} \left( \frac{3 \times 255^2}{\text{MSE}(I, \hat{I})} \right)$$

5. Conclusions

This state-of-the-art report has presented a comprehensive overview of the current research that expands LDR content for the generation of HDR images and videos. The LDR to HDR expansion methods fill the void between classic imaging and HDR imaging, allowing existing LDR content to be used for HDR applications. Even if HDR imaging has become very popular outside the computer graphics community, there are still no HDR cameras and video-cameras for the general consumer. A certain knowledge, that a general consumer is not expected to have, is needed to capture HDR images. Moreover, only capturing still images requires certain conditions such as static scenes, a tripod to avoid misalignments, and no variation in the lighting conditions. To meet all these conditions in real scenarios is very difficult and requires time and expertise.

The discussed methods for LDR to HDR expansion result in a good compromise between HDR imaging and available camera technology.

We have also shown how LDR to HDR expansion methods and related techniques can be suitable for HDR compression. These methods’ encoding procedure first uses a luminance compression stage, generally via some tone mapping operation, followed by standard LDR compression. The decoding is performed via an LDR decoding stage, followed by an expansion, potentially the inverse of the tone mapping operator used. These compression methods are particularly useful because most of them are backwards-compatible and also, as compression methods improve in the LDR fields, there is an immediate and automatic benefit for such HDR compression techniques.

The main limitations of most LDR to HDR expansion methods occurs when trying to expand large over-exposed areas. This issue clearly depends on the size of over-exposed areas in the image or video. The quality is inversely proportional to the area of over-exposed regions since large over-exposed areas imply more information to reconstruct than smaller ones. As an example of reconstruction that highlights these limitations, using Banterle et al.’s method the manual method may be detrimental for many applications. Further research in this area would be required to tackle this problem. When considering video sequences, exposure changes in between frames, showing details in over-exposed or under-exposed areas which become well-exposed may be exploited as a solution to such problems. It may be possible to project well-exposed areas from other (previous or successive) frames onto the current one where that same area is over-exposed or under-exposed.

We have presented some validation methods for identifying the quality of the expansion methods. These validation methods currently only cater for still images and IBL applications and no validation study on expanded videos yet exists. As HDR video becomes more important some form of validation for video will naturally be required. Psychophysical studies on videos with complex stimuli, such as a shot from a movie, may not be easily run, possibly automated methods may provide a more straightforward solution.
HDR technology has not yet reached its full maturity. Capturing HDR videos is still an open issue. Moreover, capturing HDR still images can be a long and tedious task when using classic LDR technology, and is still expensive with HDR cameras. The presented methods for the reconstruction of HDR content from LDR content have made the capturing of HDR content for consumers a more straightforward process. Moreover, all LDR legacy content can be utilised for HDR media, or used for HDR processing and re-lighting real and synthetic objects.

6. Acknowledgement

We thank Ahmet Oğuz Akyüz, Rafał Mantiuk, Tunç Ozan Aydin, Piotr Didyk, Paul Debevec, Greg Ward, Edward H. Adelson, Allan Rempel, and Hans van Hateren for the LDR and HDR images used in our paper. We thank Piotr Didyk and Rafał Mantiuk for their insightful comments on the draft of this STAR. Finally we thank Stanford’s Graphics Group for the Happy Buddha and Lucy models from the Stanford 3D repository.

This work reported in this STAR has formed part of EPSRC grants EP/D032148 and EP/G001634/1 whose funding and support are gratefully acknowledged.

References


[DFF04] DFF04: DALY S., FENG X. Bit-depth extension using spatiotemporal microdithering based on models of the equivalent input noise of the visual system. In Proceedings of Color Imaging VIII. Processing, Hardcopy, and Applications (June 2004), vol. 5008, SPIE.


The Eurographics Association 2009.


