High-Quality Reverse Tone Mapping for a Wide Range of Exposures

Rafael P. Kovaleski, Manuel M. Oliveira  
Instituto de Informática, UFRGS  
Porto Alegre, Brazil  
Email: {rpkovaleski,oliveira}@inf.ufrgs.br

Abstract—Reverse-tone-mapping operators (rTMOs) enhance low-dynamic-range images and videos for display on high-dynamic-range monitors. A common problem faced by previous rTMOs is the handling of under or overexposed content. Under such conditions, they may not be effective, and even cause loss and reversal of visible contrast. We present an rTMO based on cross-bilateral filtering that generates high-quality HDR images and videos for a wide range of exposures. Experiments performed using an objective image quality metric show that our approach is the only technique available that can gracefully enhance perceived details across a large range of image exposures.

Keywords—reverse tone mapping; high dynamic range; cross-bilateral filter;

I. INTRODUCTION

Recent developments in high-dynamic-range (HDR) hardware [2] indicate that HDR displays should become widely available in the near future. In this scenario, reverse tone mapping operators (rTMOs) [3], [4], [5], [6] seek to enhance the brightness and contrast of low-dynamic-range (LDR) images and videos for HDR displays. The goal of these techniques is to improve the overall viewer's experience, and their benefits have been verified by recent studies [7], [8], [9], [10].

A common limitation of previous rTMOs is their inability to handle under or overexposed content. In such situations, these operators will not improve the perceived contrast, and may even cause loss and reversal of visible contrast. We present an rTMO based on cross-bilateral filtering that generates high-quality HDR images and videos while supporting a wide range of exposure conditions. We evaluate the performance of our solution against previous rTMOs using DRIM [11], an objective image quality metric. These experiments show that ours is the only technique capable of gracefully enhancing perceived details across a large range of image exposures.

A key component of several recent rTMOs is a brightness enhancement function (BEF) [4], [5], also known as an expand map [3], [12]. BEFs are generated by determining areas where image information may have been lost (primarily due to sensor saturation) and filling these regions using a smooth function. High-quality BEFs significantly improve the result of an rTM operation [5].

Our solution is designed around the automatic and efficient creation of high-quality BEFs, and builds upon the technique proposed by Kovaleski and Oliveira [5]. While their technique properly handles under- and well-exposed images, it does not perform as well for over-exposed ones. We analyze the reasons for such behavior, and show that it results from the fact that these BEFs are computed considering, for each pixel, the maximum intensity value among the R, G, and B channels. While the authors' original motivation was to conservatively detect saturated pixels, this treats all channels equally, regardless of their perceptual differences. As a side effect, it prevents the technique from properly handling overexposed areas.

The key observation of our work follows from this analysis: the replacement of the maximum-intensity channel by the pixel’s luminance (which properly weights the perceptual contributions of the various channels) lends to a more robust solution, capable of handling a wider range of exposures. Although this difference might at first appear subtle, it is indeed quite significant: it results from a better understanding of the underlying problem, and extends the range of applicability of the technique. As a result, our solution is the only one that automatically performs reverse tone mapping of under-, well-, and over-exposed images and videos. Moreover, it leads to simpler equations that provide greater flexibility. Thus, while Kovaleski and Oliveira’s original technique only supports acceleration through the use of a bilateral grid, our solution allows one to directly use any of the existing bilateral-filtering acceleration strategies and implementations.

Figures 1 and 2 show examples of BEFs generated using our approach. Figure 1 (b) depicts a BEF created for the LDR image shown on its left. The image in (c) is a tone-mapped version of the resulting HDR image. Note the many details that were not immediately visible in the original image, which contains both under- and over-exposed areas. The image in (d) shows the result produced by DRIM, where the blue pixels represent contrast enhancement. Figure 2 shows examples of BEFs computed with our technique using various bilateral-filtering implementations. In all examples, note the ability of our solution to produce smooth functions while preserving even fine details, such as small leaves and twigs.

The contributions of our work include:

- An automatic, high-quality reverse-tone-mapping operator for images and videos that supports a wide range of exposures (Section IV). Our solution improves Kovaleski and Oliveira’s technique [5] to robustly support overexposed regions;
- An efficient approach for creating smooth high-quality...
brightness enhancement functions for rTMOs based on cross-bilateral filtering (Section IV). Our method supports any bilateral-filter acceleration technique, freeing the user from the limitations of any specific implementation.

II. RELATED WORK

Banterle et al. [3] proposed one of the first reverse-tone-mapping operators. They apply the inverse of Reinhard et al.’s tone-mapping operator [1] to an LDR image and use a median-cut algorithm to cluster light sources in areas of high luminance. A density estimation of high-luminance regions, called expand map, is used to interpolate between the input image and the initial inverse-tone-mapped image, generating the HDR output. By designing a temporally coherent expand map, this work was extended for videos [12].

The LDR2HDR framework of Rempel et al. [4] is based on the use of brightness enhancement functions, which are somehow similar to Banterle et al.’s expand maps [3]. First, the image values are linearized; then, a binary mask (containing only the pixels whose linearized values are above a certain threshold) is blurred by a 2-D Gaussian kernel and then combined with an edge stopping function to maintain the contrast on sharp edges. The resulting BEF is then used to scale the linearized image values. This strategy is computationally efficient, but sacrifices BEF quality.

Kovaleski and Oliveira [5] automatically generate high-quality edge-aware BEFs in real time using a bilateral grid [13]. Their formulation, however, does not properly handle overexposures, a condition for which HDR images are highly useful. Our work generalizes and improves Kovaleski and Oliveira’s technique, allowing it to also handle overexposures, as well as freeing it from the restriction to be used with a bilateral grid. Our solution supports all bilateral-filtering implementations and acceleration strategies.

Meylan et al. [14] split the input image into diffuse and specular components using some threshold. Specular areas are expanded and then slightly blurred, to avoid visual artifacts and unnatural contours, while diffuse areas are mildly scaled. The segmentation of bright areas was improved by Didyk et al. [15], who presented a semi-automatic technique for enhancing images and videos by labeling regions as diffuse surfaces, light sources, specular highlights, and reflections. Each area is then enhanced using different expansion functions. This system requires user intervention and is meant as a professional post-production tool.

Wang et al. [16] detail an interactive method to fill overexposed and underexposed areas with texture from other, better exposed regions, which are manually selected by the user. In contrast with this and with Meylan et al.’s technique [14], ours is completely automatic.

Masia et al. [6] performed a detailed perceptual study that considered viewers’ preferences comparing the results of three rTMOs and the original LDR image for various exposure levels. The considered rTMOs were LDR2HDR [4], Banterle et al.’s [3], and linear contrast scaling [9]. Masia and collaborators found that the performance of these operators (measured as user ratings) decrease as the number of overexposed pixels increases. The authors also found that, in the case of overexposure, scaling the images using a gamma expansion (calculated from the estimated image key [17]) is a simple yet effective technique. As the gamma-expansion approach was designed for overexposed images, it does not perform well for underexposed ones, achieving similar ratings as the other tested operators [6], and being prone to loss and reversal of visible contrast (Figures 6 and 7). More recently, Masia and Gutierrez [18] extended the work of Masia et al. [6] by proposing more robust methods to calculate the gamma value used for image expansion. These methods are still fairly
simple to compute, but show a higher correlation with the data gathered in their user studies. However, the fundamental restrictions of the original technique persist and, likewise, they are not applicable to videos. In contrast, our rTMO supports a wide range of exposures, producing high-quality results all across it, and supports reverse tone mapping of videos.

Masia et al. [21] have also proposed a higher-level approach to reverse tone mapping that considers saliency information along with pixels intensity values. It, however, depends on user input to assign area importance.

III. BILATERAL AND CROSS-BILATERAL FILTERING

A bilateral filter [22], [23], [24] is a nonlinear filter that simultaneously performs domain and range filtering. Using a Gaussian function $G$ as decreasing function, the bilateral-filtered version $I^b$ of a grayscale image $I$ is defined as:

$$I_p^b = \frac{1}{W_p^s} \sum_{q \in S} G_{\sigma_s}(|p-q|) G_{\sigma_r}(|I_p - I_q|) I_q$$

$$W_p^s = \sum_{q \in S} G_{\sigma_s}(|p-q|) G_{\sigma_r}(|I_p - I_q|),$$

where $I_p$ and $I_q$ correspond to the pixel values at positions $p$ and $q$ in image $I$, and $\sigma_s$ and $\sigma_r$ are the standard deviations of Gaussian functions for the space and range domains, respectively. $W_p^s$ is a normalization factor.

By replacing the term $G_{\sigma_r}(|I_p - I_q|)$ in Equation 1 with $G_{\sigma_r}(|E_p - E_q|)$, for a second image $E$, one obtains a cross bilateral filter [25], [26] (Equation 2). Intuitively, the Gaussian filtering weights are based on the spatial distances between pixel locations $p$ and $q$ in image $I$, and on the differences between pixel values $E_p$ and $E_q$ from image $E$.

$$I_p^b = \frac{1}{W_p^s} \sum_{q \in S} G_{\sigma_s}(|p-q|) G_{\sigma_r}(|E_p - E_q|) I_q$$

$$W_p^s = \sum_{q \in S} G_{\sigma_s}(|p-q|) G_{\sigma_r}(|E_p - E_q|).$$

Due to the nonlinearity of this filter, its use in real time applications is limited. Some techniques allow it to be computed at interactive rates by approximating the result using different methods. Durand and Dorsey [27] compute the filter response by linearly interpolating several discretized intensity values, which are filtered with a Gaussian kernel in the frequency domain. This work was extended by Paris and Durand [28] by interpreting their computation as a higher-dimensional linear convolution, which is followed by a trilinear interpolation and a division. A generalization of this idea is the bilateral grid [13]. A bilateral grid $\Gamma$ is a regularly-sampled 3-D array, where the first two dimensions define the image’s space domain, while the third one represents the image’s range. Filtering using a bilateral grid is performed in three steps: (i) grid creation, (ii) processing, and (iii) slicing. To process a bilateral grid, a 3-D function $f$ is applied to its contents, producing a new grid $\Gamma' = f(\Gamma)$. For the bilateral filter, $f$ is a convolution by a Gaussian kernel, using $\sigma_s$ and $\sigma_r$ as variances for the domain and range dimensions, respectively.

The slicing operation is used to extract a piecewise-smooth 2-D map from the bilateral grid.

Using a bilateral grid, the bilateral-filter algorithm of Paris and Durand [28], which approximates the bilateral filter in Equation 1, can be expressed as:

$$bf(I) = s_f(G_{\sigma_s,\sigma_r} \otimes g(I)),$$

where $s_f$ is the slicing operation using $I$ as reference image, $g(I)$ is the bilateral grid created using image $I$, $\otimes$ is the convolution operator, and $G$ is a 3-D Gaussian kernel with spatial and range standard deviations given by $\sigma_s$ and $\sigma_r$, respectively. With a bilateral grid, cross-bilateral filtering is performed using image $E$ to determine the grid positions, while storing the values from the data image $I$. The slicing operation then uses the edge image to recover the result.

Adams et al. [29] efficiently implement color bilateral filters in 5D using uniform simplices. The use of simplices results in cheaper interpolation operations when compared to Paris and Durand’s work [28]. More recently, Gastal and Oliveira [20] compute manifolds that adapt themselves to the signal. Weighted interpolation between these manifolds generates the output with increased accuracy and speed, using a smaller number of sampling points than previous techniques.

Fig. 3: BEF generation for a synthetic image. (left) Input image containing high R, G, and B channel values. (center) BEF generated by the technique of [5]. It treats R, G, B regions equally, despite their perceptual differences. (right) BEF generated with our solution. Note how it properly represents the subtle gradients found in (left).

IV. AN rTMO FOR A WIDE RANGE OF EXPOSURES

Kovaleski and Oliveira [5] generate brightness enhancement functions using a bilateral grid. The grid is created using two different images, much like in cross-bilateral filtering. But the slicing operation is performed with a third image. Following the same notation of Equation 3, their technique can be expressed as:

$$bf(I) = s_b(G_{\sigma_s,\sigma_r} \otimes g_A(C)),$$

where $A$, $B$, and $C$ are single-channel images. Each pixel in $A$ contains the maximum intensity value among the RGB channels of the corresponding pixel of $I$. Image $B$ contains $I$’s luminance information, and image $C$ contains all pixels in $A$ whose values are above some threshold $t$ (230 for video, 254 for images). In their formulation, the maximum-intensity value among the RGB channels (image $A$) was used as a conservative way of detecting saturated pixels. As we
Fig. 2: Brightness-enhancement functions generated with our technique using different bilateral-filter algorithms. Input image, and BEFs created using brute force, real-time bilateral filter [19], bilateral grid [13], and adaptive manifolds [20].

Fig. 4: Results of the DRIM metric [11] for the building series of bright/overexposed images. First row: input LDR images. Other rows: DRIM output for the results produced by the rTM operators of Kovaleski and Oliveira [5], Masia and Gutierrez’s [18], and ours. The colors green, blue, and red indicate pixels affected by loss of visible contrast, amplification of invisible contrast, and contrast reversal, respectively.

will show, this choice compromises the technique’s ability to properly handle overexposures. Image $I$ has its values linearized using a gamma curve of 2.2 prior to the generation of images $A$, $B$ and $C$. $g_A(C)$ is a bilateral grid filtered using the edge information from image $A$, while storing values from image $C$. $\sigma_s$ and $\sigma_r$ are set to 100 and 30, respectively.

We rewrite Equation 4 using the structure of the bilateral-filter equation (Equation 1):

$$I_p^b = \frac{1}{W_p} \sum_{q \in S} G_{\sigma_s} (\|p - q\|) G_{\sigma_r} (|B_p - A_q|) C_q$$

$$W_p^b = \sum_{q \in S} G_{\sigma_s} (\|p - q\|) G_{\sigma_r} (|B_p - A_q|),$$

where $A$, $B$ and $C$ are the same as in Equation 4.

A key observation is that the use of the maximum-intensity-
value image $A$, in Equation 5, tends to cause Kovaleski and Oliveira’s technique to be unable to enhance details in bright or overexposed LDR images. This is illustrated in Figure 3 with a synthetic image containing high R, G and B values and sharp edges. By using the maximum-intensity channel values, their technique treats the R, G, and B channels equally, despite their perceptual differences. Such inconsistencies are reflected in the resulting BEFs (Figure 3(center)). For comparison, the BEF generated with our solution is shown in Figure 3(right). Note how it properly represents the subtle gradients found in the original image.

This observation also leads to a solution to the problem: replace the maximum-intensity-channel image $A$ with the luminance image $B$. This modification not only solves the issue with overexposed images, but also greatly simplifies the BEF generation process, turning Equation 5 into a cross-bilateral filtering operation:

$$
I_b^p = \frac{1}{W_b^p} \sum_{q \in S} G_{\sigma_s}(||p - q||)G_{\sigma_r}(|B_p - B_q|)C_q
$$

$$
W_b^p = \sum_{q \in S} G_{\sigma_s}(||p - q||)G_{\sigma_r}(|B_p - B_q|).
$$

The use of Equation 6 instead of Equation 5 allows us to immediately use any bilateral-filtering acceleration technique. This is illustrated in Figure 2, which shows BEFs computed with our solution for two LDR images using four different approaches: brute-force bilateral filter (no acceleration), bilateral grid [13], real-time bilateral filter [19], and adaptive manifolds [20]. In contrast, Equation 5 would require a complete rewrite of other bilateral-filtering techniques to be adapted to the algorithm in [5]. Furthermore, we use $\sigma_s = 150$, which corresponds to an angle of 1.2° at a viewing distance of 3m for a 37” HDR display with a resolution of 1920x1080. This results in a blur filter spectrum containing low angular frequencies of 0.5 cycles per degree or less, to which the human visual system is not very sensitive [4]. We use $\sigma_r = 25$, obtained through extensive testing. In our experience, these values are better suited for BEF generation.

Given the computed BEFs, we generate the resulting HDR images using similar steps as Rempel et al. [4] and Kovaleski and Oliveira [5]: the input image $I$ is linearized using a 2.2 gamma curve. From the linearized image, we generate a BEF using Equation 6. Then, the BEF has its values scaled to the $[1, \alpha]$ range. For our experiments we use $\alpha = 4$. Finally, the linearized input image has its values scaled according to the target-display capabilities and chosen $\alpha$ value: we use 0.3 cd/m² for the black point and 1,200 cd/m² for the white point. The resulting HDR image is then obtained by pointwise multiplication of this scaled image by our rescaled BEF.
Fig. 6: Results of the DRIM metric for the car series of dark/underexposed images. First row: input LDR images. Other rows: DRIM output for the results produced by the rTM operators of Kovaleski and Oliveira [5], Masia and Gutierrez’s [18], and ours. The colors green, blue, and red indicate pixels affected by loss of visible contrast, amplification of invisible contrast, and contrast reversal, respectively.

V. RESULTS AND DISCUSSION

Masia et al. [6] compared the results of their gamma-expansion rTMO against three reverse tone mapping operators, namely the LDR2HDR [4], Banterle et al.’s operator [3], and linear contrast scaling [9]. For these comparisons, they used images taken with different exposures. They have shown that the gamma-expansion rTMO outperforms the compared techniques for overexposed images, while performing similarly for under and properly exposed images. This study, however, did not consider Kovaleski and Oliveira’s operator [5].

Masia et al. [6] calculate the gamma value using a linear regression of the form $\gamma = 10.44k - 6.282$, where $k$ is the image key [17]. In a subsequent tech report, Masia and Gutierrez [18] state that this equation was incorrectly reported and present new equations for calculating the gamma value.

To achieve a comprehensive evaluation, we compare the results of our technique against Kovaleski and Oliveira’s and the revised version of Masia et al.’s (based on the robust regression equation presented by Masia and Gutierrez [18]).

To obtain an objective evaluation of these three techniques, we use the Dynamic Range Independent Image Quality Metric (DRIM) [11]. Such metric can detect three different types of events in the resulting HDR images: loss of visible contrast (shown in green), amplification of invisible contrast (shown in blue) and reversal of visible contrast (shown in red). For a reverse-tone-mapping operator, loss and reversal of visible contrast are undesirable, while amplification of invisible contrast tends to increase perceived image quality [4].

For the experiments, we use the same images used in Masia et al.’s work [6], which are available as part of their supplemental materials. They correspond to series of under and overexposed images, some of which are shown on the top rows of Figures 4 to 7. Due to space constraints, not all images sequences are shown in the paper, but the presented results are representative of all under and overexposed sequences. The building (Figure 4) and graffiti (Figure 5) series correspond to bright or overexposed images, while the car (Figure 6) and flowers (Figure 7) series correspond to dark or underexposed images.

For the DRIM comparisons, our rTMO used a bilateral grid [13]. The results produced by other bilateral-filter implementations performed equally well with DRIM. Using the bilateral grid, our BEFs are generated in under 1ms for images up to $1920 \times 1080$ pixels on an Intel Mobile Core 2 Quad Q9000 2.0 GHz CPU, using an nVidia 9400M GPU.

Figures 4 and 5 show that for overexposed images our technique produces results that are at least as good as the ones produced by the revised version of Masia et al.’s approach [18]. Note that their technique has been designed to achieve its best performance for overexposed images. For both techniques there is an overall increase of contrast, even in areas where there is an apparent loss of detail (such as in
the lightpost in image building04 and in the graffiti in images graffiti03-04).

For overexposed images, our solution outperforms Kovaleski and Oliveira’s approach (see the larger clusters of blue pixels in the DRIM results for our technique). These results are explained by the differences between the equations that define the two methods, and by our choice of parameter values that work better for a larger range of exposures. As already discussed in Section IV, the use of the maximum-intensity-value image (image A) in Equation 5 tends to cause their technique to be unable to enhance details in bright or overexposed images. Our solution avoids these problems by properly weighting the three image channels using a luminance image (image B) in Equation 6.

Figures 6 and 7 show the results of the three rTMO applied to a series of darker/underexposed images. For the first image in both sequences, the results produced by the gamma-expansion operator [18] suffer from significant loss and reversal of visible contrast, shown in the DRIM results in green and red, respectively. For the entire series, best results are obtained with our solution, followed by Kovaleski and Oliveira’s. This should not be surprising, as the gamma-expansion technique was developed for overexposed images.

Figures 4 to 7 demonstrate the ability of our solution to handle a wide range of exposures, enhancing perceived details all across it. Thus, it outperforms the gamma-expansion operator [18], which is not appropriate for underexposed images, as well as the Kovaleski and Oliveira’s original operator [5], which is not as effective for overexposed images.

Figure 8 illustrates the use of our solution to also create HDR videos. The algorithm is independently applied to the individual frames. The top row shows frames from fire and smoke simulation videos. Their corresponding BEFs are shown in the middle row, while the result of the DRIM comparison of the original LDR and resulting HDR video frames are shown in the bottom row. Note the amplification of invisible contrast in these frames, which contain both dark and bright regions. The cost of our rTM operator is the same as the cost of the selected cross-bilateral filter implementation.

VI. CONCLUSION

We presented a high-quality reverse-tone-mapping operator for images and videos that supports a wide range of exposures. Our solution is based on the use of cross-bilateral filtering to compute smooth BEFs that preserve sharp edges. It generalizes and improves Kovaleski and Oliveira’s technique [5], and can be directly used with any bilateral-filtering strategy, freeing the user from the limitations of any specific implementation.

We performed objective comparisons of our results against the ones produced by the gamma-expansion operator of Masia and Gutierrez [18] and by Kovaleski and Oliveira’s operator [5] for a large range of image exposures using DRIM [11].